The Journal of Genetic Psychology, 2002, *156*(1), xxx-xxx

A Stage is a Stage is a Stage: A Direct Comparison of Two Scoring Systems

THEO L. DAWSON
Cognitive Science
Hampshire College

ABSTRACT. L. Kohlberg (1969) argued that his moral stages captured a developmental sequence specific to the moral domain. To explore this contention, the author compared stage assignments obtained with the Standard Issue Scoring System (A. Colby & L. Kohlberg, 1987a, 1987b) and those obtained with a generalized content-independent stage-scoring system called the Hierarchical Complexity Scoring System (T. Dawson, 2002a), on 637 moral judgment interviews (participants' ages ranged from 5 to 86 years). The correlation between stage scores produced with the 2 systems was .88. Although standard issue scoring and hierarchical complexity scoring often awarded different scores up to Kohlberg's Moral Stage 2/3, from his Moral Stage 3 onward, scores awarded with the two systems predominantly agreed. The author explores the implications for developmental research.

Key words: cognitive development, developmental assessment, developmental stages, life-span development, moral development

I EXPLORED KOHLBERG'S contention that moral stages represent a unique cognitive structure, along with broader questions about the nature of development, by comparing the functioning of two developmental stage-scoring systems—Kohlberg's domain-specific Standard Issue Scoring System (Colby & Kohlberg, 1987b), and the domain-general Hierarchical Complexity Scoring System (Dawson, 2002a).

Developmental stages, also referred to in this article as *orders of hierarchical complexity (complexity orders)*, are conceived of as a series of hierarchical integrations of knowledge structures. Most developmental stage theories use the notion of hierarchical complexity. In the Piagetian model (Piaget, 1977), for

The author thanks the Murray Research Center, Cheryl Armon, Marvin Berkowitz, and Larry Walker for the use of their moral judgment interview data. This project was funded by a grant from the Spencer Foundation. The data presented, the statements made, and the views expressed are the responsibility solely of the author.

Address correspondence to Theo L. Dawson, Cognitive Science, Hampshire College, Amherst, MA 01038; tdawson@hamshire.edu (e-mail).

example, each successive hierarchical integration produces novel understandings by using the operations of the previous order as conceptual elements in its new constructions. This notion is central to several other developmental theories as well, including those of Case (1985), Fischer (1980), and Werner (1948) and is the basis for a number of developmental scales, such as the levels and tiers of Fischer's skill theory and the stages of Commons' general stage model (Commons, Richards, with Ruf, Armstrong-Roche, & Bretzius, 1984; Commons, Trudeau, Stein, Richards, & Krause, 1998).

The notion of developmental stages leads to certain empirical expectations. First of all, developmental stages are built on each other in the sense that the construction of a subsequent stage requires the elements and operations of the previous stage. Consequently, each is logically more difficult than was its predecessor. This meant that development should proceed from one stage to the next in an invariant sequence with no skipping of stages. A large body of longitudinal evidence supports the sequentiality in the acquisition of stages of development (Armon & Dawson, 1997; Case, Okamoto, Henderson, & McKeough, 1993; Colby, Kohlberg, Gibbs, & Lieberman, 1983; Dawson, 2002b; Fischer & Bullock, 1981; Kitchener, King, Wood, & Davison, 1989; Snarey, Reimer, & Kohlberg, 1985; Walker, 1982).

However, evidence of sequentiality does not provide adequate support for the existence of developmental stages. Each stage in a hierarchical complexity sequence is generally defined by a set of internally consistent formal properties. In Piaget's model (Piaget, 1977), those stages were said to constitute a *structure d'ensemble,* or structure of the whole. In more recent formulations of developmental stages, the processes and elements of a given stage are more often considered as the processes and elements of a dynamic, complex system, and stage change is considered to be the transformation of a system of that kind into another that was more hierarchically complex (Fischer & Bidell, 1998; Stevens, 2000; van Geert, 2000). Both Piaget's and dynamic-systems theories of stage led to the notion that, at least under some conditions, development from one stage to another should appear discontinuous, with plateaus during periods of consolidation, and spurts during periods of reorganization. Fischer and Rose (1999), on the basis of a series of developmental studies, proposed one model of development as a sequence of overlapping waves, (a) with relatively long periods of consolidation (plateaus) during which performance within a domain tends to be largely homogeneous (is predominantly at a single complexity order) and (b) shorter transitional periods (spurts) characterized by vacillation between the modal complexity order and its successor. If stages exist, then valid developmental stage measurement systems should make it possible to observe spurts and plateaus in development. Fischer and Bidell (1998) have shown that under optimal conditions development from one stage to another appears wave-like. More recently Dawson, Commons, and Wilson (manuscript submitted for publication) showed that development, as measured with the Hierarchical Complexity Scoring System

(Dawson, 2002a), not only proceeds in an invariant sequence but appears wave-like, with prominent spurts and plateaus at every stage transition measured.

The notion of cognitive development as a series of hierarchical integrations is useful in psychological assessment and research only insofar as instruments can be designed to measure stages with reasonable consistency, reliability, and freedom from bias. This design has proven to be a difficult task, especially outside the logico–mathematical domain. Most of the stage-scoring systems developed to assess social, personal, or moral reasoning (Armon, 1984; Colby & Kohlberg, 1987b; Kitchener & King, 1990) are vulnerable to accusations of cultural bias because scoring with them involves matching arguments with exemplars selected from the performances of small samples of predominantly White, middle-class respondents. Moreover, there are no fixed criteria for assessing how well those scoring systems function as measures of developmental stage. Their developers do not always agree about what kind of behaviors arre evidence of the attainment of any given stage, and it is not clear how the developmental levels in one system are related to the levels in the other systems. This paper directly addresses the latter issue by asking whether a domain-based scoring system that uses a concept-matching strategy (Kohlberg's Standard Issue Scoring System; Colby & Kohlberg, 1987b) assesses the same dimension of performance as a generalized scoring system (the Hierarchical Complexity Scoring System; Dawson, 2002a), which is not dependent on concept matching.

Standard Issue Scoring System

Kohlberg's moral judgment instrument is infrequently used today. Not only did the instrument apparently fail to support postulates of stage theory, such as structured wholeness and no regressions to earlier stages (Fischer & Bidell, 1998), but it has been severely criticized for alleged cultural and gender biases (Gilligan, 1977; Puka, 1994; Simpson, 1973; Stack, 1990). In fact, as noted heretofore, most traditional stage-assessment systems suffer from bias introduced by nonrepresentative construction samples and an overdependence on particular content because, despite the incorporation of structural criteria in the stage definitions, scoring is generally a process of matching the justifications of respondents to those in a scoring manual generated from a limited number of construction cases. For example, Okonkwo (1997), in her recent study of Nigerian students' moral reasoning, noted that Kohlberg's scoring manual lacks scoring criteria for advanced forms of Nigerian moral concepts related to parental obedience, family interdependence, and the transcendental authority of a divine being. Scoring systems that rely on justification- or concept-matching introduce other problems as well. First, they can be used only to assess stage within a particular domain, which makes cross-domain comparisons difficult; and second, scoring for stage on the basis of the presence or absence of particular conceptual content makes it impossible to examine the relationship between stages and

conceptual change (Dawson, 1998; Dawson & Gabrielian, 2003).

Despite these problems, Kohlberg's Standard Issue Scoring System (Colby & Kohlberg, 1987b), is one of the best known stage-scoring systems. Kohlberg and his colleagues used what they called a *bootstrapping* process to define moral judgment stages and constructed a scoring manual. They started with a theoretical sequence strongly influenced by Piaget's (1965) idea that moral thinking moves from heteronomy to autonomy and from concrete to abstract, and refined their understanding of moral stages as they gathered successive rounds of longitudinal data from a group of boys from the New England area of the United States, the youngest of whom were 10 years old at the outset of the study. The Standard Issue Scoring System was constructed by analyzing seven sets of interviews from Kohlberg's original longitudinal study. Each of these sets of interviews included assessments from all six of the test times, which were separated by 4-yr intervals. Each performance was assigned a global-stage score "based on intensive discussion and analysis" (Colby & Kohlberg, 1987a, p. 40) using criteria from an earlier version of the scoring system. Then individual responses to each dilemma provided the basis for examples in the scoring manual.

The Standard Issue Moral Judgment Interview consists of three forms, each composed of three hypothetical moral dilemmas and from 9 to 12 standard questions per dilemma. Each question is designed to probe respondents' reasoning on at least one of the moral issues—life, law, conscience, punishment, contract, and authority. The examples provided throughout this paper were chosen from responses to the Joe interview part of Form A of the Standard Issue Scoring System. In this dilemma, Joe, a 14-year-old boy, wants to go to camp. His father promises him he can go if he saves the money for himself. So, Joe works hard at his paper route and saves up enough to go to camp. However, just before camp begins, his father changes his mind and decides he wants Joe's money to go on a fishing trip. Naturally, Joe does not want to give up going to camp, so he thinks he might not give his father the money. Follow-up questions probe for respondents' understanding of the dilemma and their reasoning about promises, ownership, and familial obligations. For example, respondents are asked if it is important for a father to keep a promise to his son, and the response is probed to reveal the reasoning behind it.

To assess moral development with the Standard Issue Scoring System, a researcher administers a set of moral judgment interviews, transcribs them, identifies each moral argument addressing one of six moral themes (life, law, conscience, punishment, authority, and contract), then uses the scoring manual to match each identified argument with a similar argument. These criterion judgments (CJs) are intended to be structural in the sense that they reflect a particular sociomoral perspective or operative level, but they are expressed in terms of the content of interviews in the construction sample. This is not to say that the Standard Issue Scoring System entirely fails to distinguish between structure and content. The six moral issues represent different conceptual categories, and the

system allows for the separate coding of moral norms and elements. Stage assignment is not dependent on which issue, element, or norm is used in an argument. However, the CJs are laden with conceptual content, and evaluators ultimately score on the basis of how well a respondent's argument matches exemplars on the same issue. This is how the notion of care came to be identified with Moral Stage 3 moral reasoning (Gilligan, 1977).

Kohlberg's moral stages describe three periods of development in the moral domain—*preconventional*, *conventional*, and *postconventional*. Each of these three periods is subdivided into two stages so that Kohlberg's model consists of six stages of moral development.

The preconventional period (Moral Stages 1 and 2) begins in early childhood and generally extends through elementary school. At Moral Stage 1, people are thought of only as particular individuals who do particular things. Actions are justified in terms of avoiding punishment and obtaining rewards. For example, the answer to the question, "Why do you do what your dad tells you?" might be "He will give me a cookie."

At Moral Stage 2, people are classified into groups according to the actions they have performed. Action is justified in terms of one's personal interests. "If I tell lies, then my friends won't play with me anymore".

The conventional period (Moral Stages 3 and 4) begins near the end of elementary school and extends across the life span of all but a small portion of the population. This period generates the conventional norms of adulthood. At Moral Stage 3, individuals are understood in terms of fixed personality characteristics, sentiments, roles, or motives. Action is justified in terms of the reputation and characterization of the individuals or groups that are involved.

At Moral Stage 4, the yardstick for evaluating the morality of an action is the preservation (or destruction) of a system—or a society. Norms, laws, rules, and regulations form a logically coherent system. Individuals are seen as complex systems of qualities that vary with circumstances. People at this stage reason in terms of how an action would affect their individual role and status within the system, as well as how it would affect the system's capability to function. The authority of societal law, individual and group rights, and duties are meaningful. "What would happen to society if everyone . . . ?" is a question characteristic of this stage.

The postconventional period (Moral Stages 5 and 6) may begin sometime after adolescence. In any known society, only a small portion of members achieves postconventional stages of reasoning. At Moral Stage 5, individuals are seen as complex systems of qualities and processes in interaction with other complex systems. People justify actions on the basis of universal abstract principles that are general in their application, irrespective of the person affected. These principles coordinate the interests of the society and the individual. Moral Stage 6 is not discussed in this article because Standard Issue Scoring System criteria for that stage have not been developed, and it was therefore not identified in the

performances in the present sample.

A recent scaling analysis of 996 moral judgment interviews scored with the Standard Issue Scoring System provided evidence for the specified developmental sequence and some evidence of structured wholeness in development from Moral Stages 3 to 5 (Dawson, 2002b). The instrument did not perform as well at the lower stages as it did at the higher stages. The failure of the Standard Issue Scoring System to adequately assess the lower stages has been addressed by several researchers. For example, Kuhn (1976), Damon (1977) argued that Kohlberg's instrument was inappropriate for assessing the moral judgment of children and that it resulted in underestimates of children's moral competence. Also, Keller, Eckensberger, and van Rosen (1989) showed that Kohlberg's Moral Stages 1 and 2 failed to account for a wide range of moral concepts expressed by young children and that it generally underestimated their ability to take the perspective of others. Dawson and Gabrielian (2003) also have expressed concern about the limitations of Kohlberg's construction sample. They argued that not only was the sample small, but the youngest respondents were only 10 years of age at the time of the initial interviews. By 10 years of age, most children from Western cultures are performing at the first level of what Fischer (1980) calls the abstract tier, which is considered analogous to Piaget's early formal operations (Ginsburg & Opper, 1988). Yet Kohlberg's Moral Stage 2 was considered analogous to Piaget's concrete operations (Colby & Kohlberg, 1987b). Fischer and Bullock (1981), Roberts (1981), and Wohlwill (1973) have all argued that it was not possible to accurately define lower stage conceptions on the basis of performances from higher stage respondents. Given the ages of his youngest respondents, that was precisely what Kohlberg and his colleagues attempted to do.

The Hierarchical Complexity Scoring System

An alternative approach to investigating the development of reasoning would be to apply a generalized method of assessment to the hierarchical complexity of performances. Several attempts of this kind have been made. Indeed, Piaget defined each of his stages in generalized terms. *Conservation*, for example, is a general feature of concrete operations and can be observed on a wide range of tasks. Case, Griffin, McKeough, and Okamoto (1992) and Fischer and Bidell (1998) have used general stage definitions extensively to scale performances across domains, but neither has disseminated a generalized scoring system. On the basis of Commons's general stage model (Commons et al., 1984; Commons et al., 1998) and Fischer's (1980) skill theory, along with elaborated conceptions of *hierarchical order of abstraction* (explained in detail later) and layers of structure (Dawson, 2001), the Hierarchical Complexity Scoring System (Dawson, 2002a) describes explicit general criteria for determining the developmental level of performance in any domain of knowledge.

Commons's general stage model and Fischer's skill theory specify 13 simi-

larly defined levels. I refer to these levels as orders of hierarchical complexity (complexity orders). The sequence (Commons name/Fischer name) is 0 = computory/single reflexive actions, 1 = sensory and motor/reflexive mappings, 2 = circular sensory–motor/reflexive systems, 3 = sensory–motor/single sensorimotor schemes, 4 = nominal/sensorimotor mappings, 5 = sentential/sensorimotor systems, 6 = preoperational/single representations, 7 = representational mappings/representational mappings, 8 = concrete/representational systems, 9 = abstract/single abstractions, 10 = formal/abstract mappings, 11 = systematic/abstract systems, and 12 = metasystematic/single principles. Fischer's level names were used for the remainder of this article.

When assessing the hierarchical complexity of a text, a rater attends to two manifestations of hierarchical complexity. The first is the hierarchical order of abstraction of the new concepts used in its arguments, and the second is the most complex logical structure of its arguments. Both are described in the following stage descriptions, which include only the six complexity orders identified in the data used in the following analyses. The examples provided in these descriptions are from Dawson and Gabrielian's (2003) analysis of the conceptions of authority and contract associated with complexity orders in a sample of 747 moral judgment interviews scored with the Hierarchical Complexity Scoring System.

At the representational mappings order, the new concepts are referred to as second-order representational sets. They coordinate or modify representational sets (the concepts constructed at the preoperational level). The popular representational mappings complexity order concept of *having favorites*, for example, could be used to rank camping and fishing. "Camping is my favorite, and fishing is my next favorite." Concepts like *being mean*, *keeping a promise*, *changing one's mind*, and *sharing* also become common at this complexity order. "[Joe's father] is just being mean; he is taking the money away from his kids." The most complex logical structure of this complexity order is linear, coordinating one aspect of two or more representations as in, "If you do not do what your father tells you to do, he will get really mad at you," in which doing what your father says and not doing what your father says are coordinated by his anticipated reaction.

At the representational systems complexity order, the new concepts are third-order representational sets. These coordinate elements of representational systems. For example, the concept of *trust*, articulated for the first time at the representational systems complexity order, could be used to describe the system of interactions between Joe and his father. "Joe trusted [his dad] that he could go to the camp if he saved enough money, and then his father just breaks it and the promise is very important." Concepts like *to turn against*, *to blame*, *to believe*, and *being fair* also are infrequently observed before this complexity order. "[If you break a promise,] they will not like you anymore, and your friends will turn against you." The most complex logical structure of this stage is multivariate, coordinating multiple aspects of two or more representations, such as, "If Joe's

Dad says Joe can go to camp, then he says he can't go to camp, that's not fair because Joe worked hard and then his Dad changed his mind," in which two conflicting representations of Dad's authority are evaluated in terms of his changed mind and Joe's hard work.

At the single abstractions complexity order, the new concepts are referred to as first-order abstractions. These coordinate representational systems. For example, the concept of *trustworthiness*, articulated for the first time at this complexity order, defines those qualities that make a person trustworthy rather than describing a particular situation in which trust is felt or not felt. It is composed of qualities that produced trust, such as telling the truth, keeping secrets, and keeping promises. "It's always nice . . . to be trustworthy. Because then, if [someone has] a secret, they can come and talk to you." Concepts such as *kindness*, *keeping your word*, *respect*, and *guilt* are also rare before the single abstractions complexity order. "If you don't do something you promise, you'll feel really guilty." The most complex logical structure of this stage identifies one aspect of a single abstraction, as in "Making a promise is giving your word" in which giving one's word is an aspect of a promise.

At the abstract mappings complexity order, the new concepts are referred to as second-order abstractions. These concepts coordinate or modify abstractions. For example, the abstract mappings concept basis could be used to coordinate the elements essential to a good relationship. "To me, [trust and respect are] the basis of a relationship, and without them you really don't have one." Concepts such as *coming to an agreement*, *making a commitment*, *building trust*, and *making compromises* also are rare before the abstract mappings complexity order. "I think [Joe and his father] could come to an agreement or compromise that they are both comfortable with." The most complex logical structure of this stage coordinates one aspect of two or more abstractions, as in "Joe has a right to go to camp because his father said he could go if he saved up the money, and Joe lived up to his commitment." Here, Joe's fulfillment of his father's conditions determines whether Joe has a right or does not have a right to go to camp.

At the abstract systems complexity order, the new concepts are referred to as third-order abstractions. These concepts coordinate elements of abstract systems. For example, the concept of *personal integrity*, which is rare before the abstract systems complexity order, refers to actions such as the coordination of and adherence to notions of fairness, trustworthiness, honesty, preservation of the golden rule, and so forth. "[You should keep your word] for your own integrity, for your own self-worth, really. Just to be the kind of person that you would want to be dealing with." Concepts such as *verbal contract*, *moral commitment*, *functional*, *development*, *social structure*, and *foundation* are also uncommon before the abstract systems complexity order. "A promise is the verbal contract, the moral commitment that the father made to his son. The only way for the child to develop his moral thinking is from watching parent's moral attitude." The most complex logical structure of this stage coordinates multiple aspects of two or more

abstractions. “Following through with his commitment and actually experiencing camp combine to promote Joe’s growth and development, not just physically but psychologically, emotionally, and spiritually.” Here multiple facets of Joe’s personal development are promoted when he keeps his commitment and accomplishes his goal.

At the single principles complexity order, the new concepts are referred to as first-order principles. These principles coordinate abstract systems. The notion of the social contract, for example, results from the coordination of human interests (where individual human beings are treated as systems). “Everybody wants to be treated equally and have a sense of fair play. Because this is so, we have an obligation to one another to enter into a social contract that optimizes equality and fairness.” Concepts such as *autonomy*, *fair play*, *heteronomy*, *higher order principle*, and *philosophical principle* are rare before the single-principles complexity order. “The only time we’re justified in breaking the social contract is when a higher principle, such as the right to life, intervenes.” The most complex logical structure of this stage identifies one aspect of a principle or axiom coordinating systems—contracts are articulations of the unique human quality of mutual trust, which coordinates human relations. Here, contracts are seen as the instantiation of a broader principle coordinating human interactions.

Dawson and her colleagues (Dawson et al., manuscript submitted for publication) used Rasch scaling[1] to investigate patterns of performance in a cross-sectional life-span sample of 747 moral-judgment interviews scored with the Hierarchical Complexity Scoring System. They found six complexity orders—representational mappings, representational systems, single abstractions, abstract mappings, abstract systems, and single principles—represented in performance between the ages of 5 and 86 years. The ages when representational-mappings to abstract-mappings complexity orders first predominate are 5, 7, 10, and 14 years. The abstract systems and single principles complexity orders do not become the plurality until 22 years of age with 3 years of college, and 26 years of age with 3 years of postgraduate work. Those age ranges were similar to those reported by Fischer and Bidell (1998) for the acquisition of analogous skill levels.

In the same paper, Dawson and her colleagues report that the sequence of acquisition of the complexity orders appears to be invariant and wavelike, with evidence of spurts and plateaus at every stage transition measured, including the transitions to the two highest complexity orders, which were identified almost exclusively in adulthood. They also demonstrate that patterns of performance in their sample are highly consistent from complexity order to complexity order. They argue that if any complexity orders in this range were incorrectly specified, or if complexity orders were left out, then much less systematic patterns of performance would be expected.

The Benefits of Domain-General Versus Domain-Specific Scoring Systems

Scoring with the Hierarchical Complexity Scoring System is different from scoring with domain-specific systems such as Kohlberg's. Most notably, the Hierarchical Complexity Scoring System does not involve matching performances to examples in a scoring manual. Instead, the rater determines the complexity order of a performance by looking for the highest hierarchical order of abstraction evident in its elements and then examines the logical structures coordinating those elements. Consequently, unlike the Standard Issue Scoring System and other systems like it, which can be used reliably only when concepts in a given performance can be matched to those in a manual, the Hierarchical Complexity Scoring System could theoretically be used to score any performance. It also could be used to score performances in any domain, using identical criteria across domains. Kohlberg and his colleagues attempted to distinguish between structure and content in their stage definitions; however, their scoring system relies on a concept-matching strategy. Consequently, it is difficult to ask questions about the relationship between structure and content except in the limited sense of asking how moral stage is related to performance on a given issue, element, or norm. The Hierarchical Complexity Scoring System takes the separation of structure and content a step further. When scoring with the Hierarchical Complexity Scoring System, particular conceptual content is the focus only to the extent that it reveals the hierarchical order of abstraction of a performance; therefore, complexity-order assignment and conceptual analyses can be conducted separately. Separate analyses make it possible to address questions about the relationship between meaning and development, which can not legitimately be addressed when researchers use scoring systems that involve concept matching.

For example, by using the Hierarchical Complexity Scoring System—along with an independently conducted content analysis—to score moral judgment interviews, one can directly address the question of whether, at a given complexity order, girls refer more to care and boys refer more to justice in their moral justifications. In other words, one can distinguish between hierarchical development and other influences on performance.

A developmental scoring system like the Hierarchical Complexity Scoring System, which uses scoring criteria that are independent of particular content, confers at least two other advantages, particularly in sociomoral domains of knowledge, where problems are ill-structured and often do not have correct answers. First, unlike domain-based stage scoring systems, the Hierarchical Complexity Scoring System does not incorporate normative statements about the conceptual content of performances. The claim that one complexity order is higher than another simply means that it is more hierarchically complex. Second, the time and expense of producing a different scoring system for every domain of knowledge are not necessary, and the need for raters to go through arduous learning processes for different systems is eliminated.

Do Moral Stages Require Their Own Assessment System?

The considerable differences between the Hierarchical Complexity Scoring System and the Standard Issue Scoring System raise the question of whether these systems measure the same dimension of performance. One way of addressing thisquestion is to determine whether the Hierarchical Complexity Scoring System and the Standard Issue Scoring System provide the same stage assessments. Kohlberg would not have expected this to be the case. Kohlberg, along with other proponents of domain theory (Cosmides & Tooby, 1994; Demetriou & Efklides, 1994), argued that development in different knowledge domains involves fundamentally different processes. For Kohlberg, the moral domain constitutes a unique structure parallel to, but progressively differentiated from, both the Piagetian logico–mathematical domain and the perspective-taking domain. In fact, Kohlberg (Colby & Kohlberg, 1987a; Kohlberg & Candee, 1984) proposed that the parallel stages of each are necessary, but not sufficient, for a given stage of moral judgment. He argued that perspective-taking development involves capacities in addition to basic cognitive capacities. Cognitive development regards the objective environment, whereas perspective-taking involves comprehension of how people think and act toward each other. Moral development demands an additional step—knowledge of how people should think and act toward one another. For example, Piaget's stage of consolidated formal operations (analogous to the abstract-systems complexity order) enables individuals to understand and manipulate systems of variables. This ability provides a foundation for attaining a level of role-taking at which individuals can understand themselves and others in terms of each person's place in the system. Finally, on the basis of the knowledge of a person's position in the system, an individual can make prescriptions of fairness or justice.

Perhaps Kohlberg would have regarded a scoring system that examined only the hierarchical order of abstraction and logical structure of moral judgment performances (without taking into account role-taking or moral concepts per se) as a method of assessing its logico–mathematical stage rather than as a method of assessing its stage of moral reasoning. Given this, if Kohlberg's notion of moral stages as unique structures associated with logico-mathematical structures in a necessary but not sufficient relationship is correct, one might expect individuals to invariably achieve a given complexity order prior to achieving an analogous moral stage.

Early evidence in support of the necessary but not sufficient hypothesis was less than compelling. According to Colby, Kohlberg, and Kauffman (1987), the strongest evidence for the radical claim of necessary but not sufficient domain structures came from intervention studies that attempted to train moral reasoning. They claimed that the best such study was conducted by Walker (1980; see also Arbuthnot, Sparling, Faust, & Kee, 1983; Faust & Arbuthnot, 1978).

Walker (1980) evaluated children with six tasks assessing three levels of cog-

nitive development (beginning operations, concrete operations, and basic formal operations), Selman's perspective-taking interview, and Kohlberg's Standard Issue Scoring System. General cognitive development and perspective taking were hypothesized to be prerequisites for moral reasoning. Because the only children who demonstrating evidence of Moral Stage 3 had fulfilled both prerequisites, Walker concluded that cognitive and perspective-taking development were "necessary but not sufficient conditions for moral development" (Walker, p. 137). However, Walker's results should be interpreted with caution. Only performances at beginning formal operations, concrete operations, Moral Stage 2/3 perspective-taking, and Moral Stage 2 were examined. All others, including transitional performances, were excluded. At best, Walker's analysis suggests that moral development lagged behind cognitive and perspective-taking development at the juncture between concrete and formal operations. Beyond supporting a developmental lag hypothesis, that study provides no additional evidence supporting the necessary or sufficient claim. Also, specifically for Walker's study (but see also Arbuthnot et al., 1983; Faust & Arbuthnot, 1978), any generalizations regarding the complete stage theory are inappropriate on the basis of evidence from one stage transition.

Four earlier reports addressed the question of whether complexity orders, as identified with the Hierarchical Complexity Scoring System, represent the same latent dimension of performance assessed with the Standard Issue Scoring System. In the first of these (Dawson, 2001), the scoring behavior of five raters trained in the Hierarchical Complexity Scoring System was compared with the scoring behavior of three raters trained in Kohlberg's Standard Issue Scoring System. All raters scored the same 43 texts. A mean score for each text was calculated for each group of raters, resulting in two scores for each text—one on the basis of the ratings of Hierarchical Complexity Scoring System raters and one on the basis of the ratings of Standard Issue Scoring System raters. Despite the fact that the two groups of raters used different scoring criteria, these mean scores were within one complexity order of one another 95% of the time (r = .94).

In a second study, Dawson (2003c) compared performances on the Hierarchical Complexity Scoring System, Standard Issue Scoring System, and Good Life Scoring System (Armon, 1984). These scoring systems were applied to different data—good education interviews, moral judgment interviews, and good life interviews, respectively—collected from the same respondents. The authors reported a correlation of .90 between moral judgment (scored with the Standard Issue Scoring System) and good education (scored with the Hierarchical Complexity Scoring System) performances, and a correlation of .92 between good life reasoning (scored with the Good Life Scoring System) and good education (scored with the Hierarchical Complexity Scoring System). Dawson argued that those correlations, combined with patterns in the acquisition of analogous good life stages, moral stages, and complexity orders provided evidence that the three scoring systems predominantly assess the same latent dimension—hierarchical

complexity. Moreover, patterns of performance on good life stages and complexity orders were more stage-like than were patterns of performance on the Standard Issue Scoring System, which suggested that Kohlberg's stages were not as consistently specified as were complexity orders and good life stages.

In a third study, Dawson, Xie, and Wilson (2003) conducted a multidimensional partial credit Rasch analysis of the relationship between scores obtained with the Standard Issue Scoring System and scores obtained with the Hierarchical Complexity Scoring System on 378 moral judgment interviews from respondents aged 5 to 85 years. They found a correlation of .92 between scores awarded with the two scoring systems, which suggested that those two systems assessed the same dimension of performance, though the Hierarchical Complexity Scoring System appeared to be somewhat easier (i.e., award somewhat higher scores) than the Standard Issue Scoring System, particularly at the lower complexity orders. The Hierarchical Complexity Scoring System also produced more stage-like patterns of performance than did the Standard Issue Scoring System.

In a final study, Dawson and Gabrielian (2003) examined the relationship between Kohlbergian moral stages and the conceptual content associated with complexity orders. They found strong correspondences between Moral Stages 2/3, 3, 4, and 5 and the conceptual content associated with the single abstractions, abstract mappings, abstract systems, and single principles complexity orders. However, the correspondences between Moral Stages 1 and 2 and the representational mappings and representational systems complexity orders were less straightforward. Some concepts associated with Moral Stage 2 were found in the representational mappings performances of 5-year-olds, whereas some concepts associated with Moral Stage 1 were found for the first time in the representational systems performances of 7- and 8-year-olds.

In the following analysis, scoring patterns within stages are closely examined to determine precisely where and how stage scores on the Standard Issue Scoring System and Hierarchical Complexity Scoring System agree and differ.

Method

It is difficult to specify a relationship between Kohlbergian moral stages and complexity orders. I approached that problem from two perspectives. Here, I approached this relationship between Standard issue and hierarchical complexity coding criteria by scoring—with the Hierarchical Complexity Scoring System—the 219 criterion judgments from Form A of the Standard Issue Scoring System. Second, I tested these relationships by comparing scoring results in a set of 637 moral judgment interviews.

Analysis 1

Form A of the Standard Issue Scoring Manual is composed of 219 criterion

judgments (CJs). Each of those provide a general description of conceptualizations at a given stage on a particular issue along with exemplars primarily drawn from Kohlberg's seven construction cases. Two raters (the author and a graduate student originally trained in hierarchical complexity scoring by the author) independently scored each of those CJs with the Hierarchical Complexity Scoring System. The initial rate of agreement was 80% within one half of a complexity order. Where disagreements occurred, the two raters discussed complexity order assignments until they came to a consensus, with the goal of assigning each CJ to a single complexity order. Most CJs were assigned to a single complexity order. However, some CJs combined qualities of two adjacent complexity orders. Those received a combined score—representational mappings/representational systems, representational systems/single abstractions, single abstractions/abstract mappings, abstract mappings/abstract systems, or abstract systems/single principles.

Analysis 2

For the second analysis, 637 interviews were collected from five Kohlbergian studies conducted between 1955 and 2000 (Armon & Dawson, 1997; Berkowitz, Guerra, & Nucci, 1991; Colby et al., 1983; Walker, 1989; Walker, Gustafson, & Hennig, 2001). The sample was distributed as shown in Table 1. The diverse population represented a wide range of socioeconomic and ethnic groups from the Berkowitz sample of children from working-class backgrounds and their parents to the Kohlberg sample of boys. However, it was not possible to report a consistent account of socioeconomic status and ethnicity because discrepant reporting methods were used in the various studies. The researchers conducted interviews with 379 (59.5%) males and 258 (40.5%) females who were included in the sample. All interviews were recorded and transcribed as specified in the Standard Issue Scoring Manual. Respondents received only Form A or Form B, or Forms A, B, and C of the moral judgment interview, depending on the study in which they participated.

All of the interviews were scored by the original research teams using the specifications in Kohlberg's Standard Issue Scoring Manual (Colby & Kohlberg, 1987b). The system assigns stage scores in substage increments across six moral issues—life, law, conscience, punishment, contract, and authority. First, each time an issue was identified in a performance, it was given a stage score. Then, a weighted average score (weighted toward the higher stage responses) for each issue was calculated from the accumulated stage scores on each issue as described in Colby and Kohlberg (1987a). Unfortunately, these weighted average scores obscured some of the variation within performances, but it was not possible to obtain the raw scores for most of the data sets. Full- and substages 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0 were identified for each of the moral issues. Age statistics for each stage and substage are shown in Table 2.

Colby and Kohlberg (1987b) concluded that standard issue scores were reli-

able within aproximately one third of a stage, with interrater correlations of .92 to .98, and alternate form correlations of .83 to .98.

I conducted hierarchical complexity scoring as described in the Hierarchical Complexity Scoring System Manual (Dawson, 2002a). The text segments (protocols) scored by the hierarchical complexity rater were responses of up to 18 standard probe questions posed by interviewers administering Form A (the Heinz and Joe dilemmas) of the moral judgment interview. Consequently, the scoring unit for the Hierarchical Complexity Scoring System was different from the scoring unit for the Standard Issue Scoring System in several ways. First, only the Joe and Heinz interviews were scored with the Hierarchical Complexity Scoring System. Second, the scoring was not issue scoring. Rather, the scoring units were the complete responses to standard probes, with one score awarded for each response. Third, no transitional scores were awarded. If a performance appeared to be transitional between complexity orders, the higher complexity order score was awarded.

Scoring with the Hierarchical Complexity Scoring System involves identifying both the highest hierarchical order of abstraction and most complex form of logic in text performances. A protocol is considered to be at a given complexity order if its elements embody the hierarchical order of abstraction of that complexity order, and the complexity of its logical structure meet the formal requirements of that complexity order. For example, a child might say, “It is worse for a father to break a promise [than a son] because he is older and knows not to lie.” The order of abstraction here is second order representations—promise and lie.[2] The logical structure is a representational system—if father is both older and knows not to lie, then it is worse for him to break a promise than it is for a son who is younger and may not know not to lie.

In the interviews, six complexity orders were identified: representational mappings, representational systems, single abstractions, abstract mappings, abstract systems, and single principles. One complexity order score was awarded to each protocol. Ideally, a protocol should represent a complete argument on a given topic. Fragmentary arguments are usually treated as unable to be scored because they tended to be down-scored. However, because this results in a loss of data, fragmentary protocols are scored if adjacent protocols in a given text provided enough information to aid in their interpretation. That means the rater had access to the entire interview when scoring. That is the standard practice in that type of research (Armon, 1984; Colby & Kohlberg, 1987a). For each case, a mean score was calculated from the protocol scores. Age statistics for each complexity order and transition are shown in Table 2.

Pearson product–moment correlations among scores of four independent raters on a subset of 112 randomly selected cases ranged from .95 to .98. Agreement rates ranged from 80% to 97% within half a complexity order and from 98% to 100% within a full complexity order. That equals or exceeds interrater agreements commonly reported in the field (Armon, 1984; Colby & Kohlberg, 1987a).

Results

Analysis 1

The overall Pearson product–moment correlation between the complexity orders assigned to CJs and the standard issue stages to which CJs were assigned was .97, indicating that the ordering of CJs specified in standard issue scoring was virtually identical to their complexity ordering. The following analysis examins the relationship in more detail by breaking down the results by complexity order. These results also are summarized in Table 3.

The representational mappings complexity order corresponds predominantly to Kohlberg's Moral Stage 1, though the fit is far from ideal. Of the 13 CJs from Form A of the Standard Issue Scoring System assigned to Kohlberg's Moral Stage 1, six (46%) were assigned to the representational mappings complexity order, four (31%) included elements of both representational mappings and representational systems complexity orders, and three (23%) were assigned to the representational systems complexity order. Clearly, standard issue criteria for Moral Stage 1 and hierarchical complexity criteria for the representational mappings complexity order are not an ideal match. Given the direction of the differences, one would have expected standard issue scores awarded to moral judgment performances to be somewhat lower than hierarchical complexity scores at that level.

The representational systems complexity order corresponded predominantly to Kohlberg's Moral Stage 2. Out of the 25 CJs from Form A of the Standard Issue Scoring System assigned to Kohlberg's Moral Stage 2, two (8%) included elements of both representational mappings and representational systems complexity orders, ten (40%) were assigned to the representational systems complexity order, eight (32%) included elements of both representational systems and single abstractions complexity orders, and five (20%) were assigned to the single abstractions complexity order. Standard issue scoring criteria for Moral Stage 2 and hierarchical complexity scoring criteria for the representational-systems complexity order are not a good match. More than 50% of the CJs for Moral Stage 2 were scored higher than the representational systems complexity order. Given the direction of the differences, one would expect standard issue scores to be lower than hierarchical complexity scores in the range of the representational systems order.

The single abstractions complexity order corresponds predominantly to Kohlberg's transitional Moral Stage 2/3. Out of the 17 CJs from Form A of the Standard Issue Scoring System assigned to transitional Moral Stage 2/3, 16 (94%) were assigned to the single abstractions complexity order, and one (6%) was assigned to the abstract mappings complexity order. The match between the hierarchical complexity and standard issue scoring criteria is excellent at the single abstractions order and one would expect a good match between hierarchical com-

plexity and standard issue scores awarded at that level.

The abstract mappings complexity order corresponds predominantly to Kohlberg's Moral Stage 3. Out of the 40 CJs from Forms A and B of the Standard Issue Scoring System assigned to Moral Stage 3, two (5%) were assigned to the single abstractions complexity order, five (18%) had elements of both the single abstractions and abstract-mappings complexity orders, and 33 (83%) were assigned to the abstract mappings complexity order. The match between hierarchical complexity and standard issue scoring criteria is good at the abstract mappings order, thus one would expect a good match between hierarchical complexity and standard issue scores obtained at that level.

The abstract systems complexity order corresponds predominantly to Kohlberg's Moral Stage 4. Out of the 46 CJs from Form A of the Standard Issue Scoring System assigned to Moral Stage 4, 1 (2%) included elements of both the abstract-mappings and abstract systems complexity orders, 33 (72%) were assigned to the abstract systems complexity order, and 12 (26%) included elements of both the abstract systems and single principles complexity orders. The match between hierarchical complexity and standard issue scoring criteria is good at the abstract systems order and one might expect a good match between hierarchical complexity and standard issue scores obtained at this level.

The single principles complexity order corresponds predominantly to Kohlberg's Moral Stage 5. Out of the 20 CJs from Form A of the Standard Issue Scoring System assigned to Moral Stage 5, all 20 were assigned to the single principles complexity order. The match between hierarchical complexity and standard issue scoring criteria is excellent at the single principles order and one would expect an excellent match between hierarchical complexity and standard issue scores obtained at that level.

Above the representational systems complexity order, the results of the direct comparison of scoring criteria suggests that a strong correspondence between scores obtained with standard issue and hierarchical complexity scoring might be expected. However, standard issue and hierarchical complexity scoring differ in several ways that might have an impact on the actual scores awarded with each system. First, standard issue scores are based on scores awarded on six moral issues. An argument is considered scorable if it is a justification for a moral position and is awarded a score on the basis of a close match to a criterion judgment in the Standard Issue Scoring System manual (Colby & Kohlberg, 1987b). Within issues, the scores are averaged, after which a grand mean of the the mean scores is taken to obtain a summary score. These averages are weighted toward the higher levels attained by the respondent, which further complicates matters. Hierarchical complexity scores are also averages; however, each score included in these averages were based on the entire response to one of 18 standard probe questions—the complete argument, including conceptual material not included in the Standard Issue Scoring System manual. A text segment is considered scorable if it constituted a complete argument or set of arguments on a topic and is scored

at its highest observed complexity order. Hierarchical complexity scores are not weighted. I hypothesize, on the basis of these observations and on my experiences working with the two scoring systems, that the hierarchical complexity approach to scoring may produce scores that are somewhat higher (within the range of approximately one half of a complexity order) than are those produced with standard issue scoring because they are based on larger text segments and because the highest complexity order is awarded, but the standard issue scoring practice of weighted averaging might counter that trend.

Taking into account both the analysis of standard issue CJs and observations about the scoring protocols of the Standard Issue Scoring System and Hierarchical Complexity Scoring System, I expect that below the single abstractions complexity order, the Hierarchical Complexity Scoring System will award scores up to a full complexity order higher than those at the Standard Issue Scoring System. However, from the single abstractions complexity order on, I expect the differences between the Standard Issue Scoring System and the Hierarchical Complexity Scoring System to introduce random variation or noise to the analysis, with a possible tendency for the Hierarchical Complexity Scoring System to award slightly higher scores. However, if the Hierarchical Complexity Scoring System and Standard Issue Scoring System assess the same dimension of performance, and the stages that are theoretically analogous are also empirically analogous, then I expect to find means for complexity orders that are reasonably close to analogous Standard Issue Scoring System means.

Analysis 2

The Pearson product–moment correlation between mean scores obtained with standard issue scoring and hierarchical complexity scoring was .88, comparable with interrater and alternate form correlations reported for the Standard Issue Scoring System. Figure 1 shows the relationship between mean standard issue scores and mean hierarchical complexity scores, with confidence intervals for the means. As anticipated, the mean complexity orders for Kohlbergian Moral Stages 3, 4, and 5 were abstract mappings, abstract systems, and single principles, respectively. The location of the representational systems complexity order mean met expectations as well. It is two thirds of a complexity order above the theorized intersection of the representational-systems complexity order and Kohlberg's Moral Stage 2. However, the mean complexity order at Kohlberg's 2/3 transitional stage, which was expected to be at single abstractions, is about one third of a complexity order higher. I did not anticipate that difference.

Overall, 84% of the complexity orders matched Kohlberg's moral stage scores within one half of a Kohlbergian stage. Table 4 shows the mean agreement rates between the standard issue and hierarchical complexity scores, broken down by moral stage. Agreement rates are high at Moral Stages 3, 4, and 5 and lower at Moral Stages 2 and 2/3. Unfortunately, because few protocols and no perfor-

mances were scored at Moral Stage 1 in this sample, I could not test the relationship between the representational mappings complexity order and Moral Stage 1.

The appropriate locations (with respect to the Standard Issue Scoring System) of the means for the abstract mappings, abstract systems, and single principles complexity orders, combined with the high rates of agreement between hierarchical complexity and standard issue scores in that range mostly confirm that the Standard Issue Scoring System and Hierarchical Complexity Scoring System measure the same dimension of performance in that range. However, the relatively large difference between scores obtained with standard issue and hierarchical complexity scoring at the representational-systems and single-abstractions complexity orders suggest that these scoring systems do not work well together in that range. The clear bias toward higher hierarchical complexity scores rather than standard issue scores suggests that the difference between the two scoring systems is not necessarily in the developmental sequence itself. In other words, the two systems may measure the same developmental sequence, but the Hierarchical Complexity Scoring System is easier (awards higher scores) than is the Standard Issue Scoring System in that range.

Discussion

There are at least two possibilities suggested by my findings. First, part of Kohlberg's Standard Issue Scoring System may assess a dimension of performance that is different from that assessed with the Hierarchical Complexity Scoring System, such that at the lower stages the stage of a moral argument might lag behind its hierarchical complexity. The necessary but not sufficient relationship Kohlberg hypothesized to exist between logico-mathematical reasoning and moral reasoning might be used to explain that lag. However, that relationship breaks down at the higher levels, where mean hierarchical complexity scores match mean standard issue scores for analogous levels. If moral stages represent a dimension of performance distinct from hierarchical complexity, one would expect to see some kind of systematic difference across all of the levels.

A second possibility is that the lower levels in either the Hierarchical Complexity Scoring System or the Standard Issue Scoring System are not as well specified as the higher levels. As previously discussed, there is reason to question the validity of the lower stages of the Standard Issue Scoring System. The system was constructed by analyzing the interviews of seven respondents who were assessed six times at 4-yr intervals. The youngest of these respondents was 10 years old at the time of the first test. Ten years of age is the median age for single abstractions on the Hierarchical Complexity Scoring System and the age at which single abstractions appear in optimal level assessments, according to Fischer and Bidell (1998). Clearly, standard issue Moral Stage 2 CJs are based largely on material from single abstractions or transitional to single abstractions performances. The empirical findings are consistent with that interpretation. First, a

large percentage (54%) of Moral Stage 1 CJs have representational systems elements, and a large percentage (52%) of Moral Stage 2 CJs have single abstractions elements. Second, the ages for attainment of Moral Stages 2 and 2.5 are considerably higher than were the ages for the acquisition of analogous levels in other stage-scoring systems, including the levels and tiers of Fischer's skill theory (1980), whereas the ages for the attainment of the representational systems and single abstractions complexity orders match those commonly reported in the literature.

A third and final source of evidence in support of the conclusion that the Standard Issue Scoring System is a less consistent measure of stage, particularly at the lower levels, than is the Hierarchical Complexity Scoring System is presented in Figure 2. In this figure, the y-axis represents the range of scores across moral issues (or protocols) within cases. For example, the figure shows that the range of issue scores for an average performance scored at Moral Stage 2 was one full stage. Likewise, the range of protocol scores for an average performance scored at the representational-systems complexity order was .35 of a stage. Moral stages (top) and complexity orders (bottom) are shown on the x-axis. Within cases, the range of moral issue scores is greater at the lower stages than it is at the higher stages ($r = -.18$, $p < .01$), indicating either (a) greater variation in performance at the lower stages, or (b) less accurate measurement at the lower stages. Moreover, the range of scores is always larger for moral stages than it is for complexity orders. Patterns of performance on complexity orders are dramatically different than they are on moral stages. First, the average amount of variation within performances is relatively stable across complexity orders. Second, there was a jagged pattern that shows greater variation during transitions than at full complexity orders and it provides evidence of consolidation at full stages and stage mixture during transitions. That pattern is consistent with other evidence for consolidation and transition on interview performances scored with the Hierarchical Complexity Scoring System (Dawson, 1998; Dawson, in press; Dawson et al., manuscript submitted for publication). Interestingly, it also is consistent with Kohlberg's (1969) original expectation that his scoring system would reveal evidence of structured wholeness (Piaget's structure d'ensemble).

Conclusion

The Standard Issue Scoring System and Hierarchical Complexity Scoring System, above the single abstractions complexity order and within the bounds of error, provide equivalent stage estimates. The correlation of .88 between mean standard issue and hierarchical complexity scores is comparable with Standard Issue Scoring System interform and interrater correlations reported by Colby and Kohlberg (1987a). Further, at the higher stages, within-stage agreement rates within one half of a Kohlbergian stage are as high as commonly reported interrater or alternate form agreement rates for the Standard Issue Scoring System

(84% to 89%). Below the single abstractions complexity order, the evidence produced here and reported in previous research indicates that the Standard Issue Scoring System functions poorly. The Hierarchical Complexity Scoring System performs well from the representational mappings complexity order to the single principles complexity order, with strong evidence of invariant sequence and developmental spurts and plateaus that are consistent with expectations for developmental stages. The cumulative evidence indicates that a domain-independent developmental scoring system can be used to measure stages of development as well as, or better than, domain-dependent systems.

These results, along with other evidence I presented, indicates that the Hierarchical Complexity Scoring System and the Standard Issue Scoring System assess the same latent dimension of performance—hierarchical complexity. In other words, the Standard Issue Scoring System does not seem to measure the development of a distinct moral structure. Rather, to the extent that it assesses developmental stage, it is a measure of hierarchical development in the moral domain.

This finding has important implications for developmental theory. First, it provides additional support for general stages of development on the basis of the notion of hierarchical complexity. Case and Okamoto (1996), Fischer and Bidell (1998), Halford (1999), and Pascual-Leone and Goodman (1979) argued that development includes general processes as well as processes tied to particular domains and contexts. These general processes involve increases in the hierarchical complexity of thought and permit the specification of general stage-scoring criteria.

The second implication for developmental theory is methodological. The importance of understanding the relationship between form and content in the study of conceptual development is widely recognized (Kuhn, 2000; Levine, 1979; Miller, 1988; Overton, Ward, Noveck, & Black, 1987). I do not suggest that form and content represent natural categories, but I find it useful to differentiate between these two aspects of performance. When a rater first assesses the complexity order of performances and then examines its conceptual content, questions can be addressed about the interrelationship that cannot otherwise be addressed. For example, one can ask about differences between the developing moral conceptions of girls and boys or examine similarities and differences in the specific conceptual content of performances across domains, culture, and context. This type of approach has already been used by Case, Okamoto, Henderson, and McKeough (1993) and Fischer and Silvern (1985). Unfortunately, many comparisons of reasoning across groups ignore the hierarchical complexity dimension, use age as a proxy for development, or simply assume that domain-based scoring systems function without bias across groups, cultures, and contexts (Adler, 1989). In such research, stage-developmental and other effects could have easily been conflated.

The use of generalized stage assessments make it possible to distinguish

between the form of thought and its content at a deep level of structure, one layer below the level of structure represented by Kohlberg's sociomoral perspective. In fact, sociomoral perspective can be viewed as content with respect to complexity order; it is a manifestation of even more abstract logical relations. Piaget (1985) described this phenomenon, in which elements of thought can be considered as content or structure dependent on the referent.

A reliable domain-independent stage-scoring system could be used widely to address questions about stages and stage transitions because it permits differentiation between form and content at a deep level. Measurement problems significantly hamper some aspects of developmental research. For example, different types of developmental scoring systems lead to different conclusions about developmental processes. Some systems produce results that support a smooth, cumulative model of learning and development; however, the Hierarchical Complexity Scoring System produce results that support a notion of learning and development that involves a series of qualitative changes. Scaling models of stage assessments made with the Hierarchical Complexity Scoring System produce compelling evidence for stage-like developmental progress, including invariant sequence and developmental spurts and plateaus (Dawson et al., in review). Scaling models of stage assessments made with the Standard Issue Scoring System have not provided compelling evidence for developmental spurts and plateaus, even though they have produced evidence for invariant sequence (Dawson, 2002c). The first result supports a dynamic systems model of development and the other does not. I argue that developmental stages are best thought of as orders of hierarchical complexity and cannot be measured accurately with domain-dependent measurement systems. Such systems contaminate stage assessments by tying stage definitions to particular conceptual content that are invariably derived from limited samples of reasoning in a given domain. This contamination adds noise to assessments that obscures aspects of developmental processes, such as developmental spurts and plateaus.

Though I provided data for some of the equivalences and differences between hierarchical complexity and standard issue stage assessments, that evidence is not adequate for a complete understanding of their relationship. The results of my research should be viewed within the context of a series of scaling studies of those and other developmental assessment systems (Dawson, 1998; Dawson, 2000; Dawson, 2002c; Dawson et al., manuscript submitted for publication) as well as an examination of the relationship between the conceptual content of Kohlberg's moral stages and the moral concepts associated with orders of hierarchical complexity (Dawson & Gabrielian, 2003). In keeping with Overton's (1998) description of a relational methodology, each of these projects approaches questions about conceptual development from a different perspective and with a different set of analytical tools. The result is an emerging account of the interconnections among general developmental processes and conceptual development in the moral domain.

NOTES

1. Rasch scaling is increasingly used to examine sequence, unidimensionality, and other aspects of the construct validity of developmental scales. See Bond and Fox (2001) for an accessible introduction to the model.

2. *Promise* and *lie* are second order representations because their underlying concepts constitute arguments about relations between first order representations. For example, for a child to construct the notion of a lie as an intentional untruth, the child has to coordinate conceptions of *true* vs. *not true* with intention. Notions of truth and intention (on purpose) appear for the first time at the single-representations complexity order.

REFERENCES

Adler, L. L. (Ed.). (1989). *Cross-cultural research in human development: Life span perspectives*. New York: Praeger.

Arbuthnot, J., Sparling, Y., Faust, D., & Kee, W. (1983). Logical and moral development in pre-adolescent children. *Psychological Reports, 52,* 209–210.

Armon, C. (1984). *Ideals of the good life: Evaluative reasoning in children and adults.* Unpublished doctoral dissertation, Harvard University, Boston.

Armon, C., & Dawson, T. L. (1997). Developmental trajectories in moral reasoning across the life span. *Journal of Moral Education, 26,* 433–453.

Berkowitz, M. W., Guerra, N., & Nucci, L. (1991). Sociomoral development and drug and alcohol abuse. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (pp. 35–53). Hillsdale, NJ: Erlbaum.

Bond, T., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement for the human sciences.* Mahwah, NJ: Erlbaum.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.

Case, R., Griffin, S., McKeough, A., & Okamoto, Y. (1992). Parallels in the development of children's social, numerical, and spatial thought. In R. Case (Ed.), *The mind's staircase* (pp. 269–284). Mahwah, NJ: Erlbaum.

Case, R., & Okamoto, Y. (1996). *The role of central conceptual structures in the development of children's thought* (Vol. 61, pp. 1–2). New York: Blackwell.

Case, R., Okamoto, Y., Henderson, B., & McKeough, A. (1993). Individual variability and consistency in cognitive development: New evidence for the existence of central conceptual structures. In R. Case & W. Edelstein (Eds.), *The new structuralism in cognitive development: Theory and research on individual pathways* (pp. 71–100). Basel, Switzerland: S. Karger.

Colby, A., & Kohlberg, L. (1987a). *The measurement of moral judgment: Vol. 1. Theoretical foundations and research validation*. New York: Cambridge University Press.

Colby, A., & Kohlberg, L. (1987b). *The measurement of moral judgment: Vol. 2. Standard Issue Scoring Manual*. New York: Cambridge University Press.

Colby, A., Kohlberg, L., Gibbs, J., & Lieberman, M. (Eds.). (1983). *A longitudinal study of moral judgment* (Vol. 48). Monographs of the Society for Research in Child Development New York: Blackwell.

Colby, A., & Kohlberg, L. (1987a). Theoretical introduction to the measurement of moral judgment. In *The measurement of moral judgment: Vol. 1. Theoretical foundations and research validation* (pp. 1–35). New York: Cambridge University Press.

Commons, M. L., Richards, F. A., with Ruf, F. J., Armstrong-Roche, M., & Bretzius, S. (1984). A general model of stage theory. In M. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations* (pp. 120–140). New York: Praeger.

Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, S. A., & Krause, S. R. (1998).

Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review, 18,* 237–278.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). New York: Cambridge University Press.

Damon, W. (1977). Measurement and social development. *Counseling Psychologist, 6,* 13–15.

Dawson, T. L. (1998). *"A good education is . . . " A life-span investigation of developmental and conceptual features of evaluative reasoning about education.* Unpublished doctoral dissertation, University of California at Berkeley.

Dawson, T. L. (2000). Moral reasoning and evaluative reasoning about the good life. *Journal of Applied Measurement, 1*(4), 372–397.

Dawson, T. L. (2001). Layers of structure: A comparison of two approaches to developmental assessment. *Genetic Epistemologist, 29,* 1–10.

Dawson, T. L. (2002a). *The Hierarchical Complexity Scoring System*. Retrieved November, 2002, from http://gseacademic.harvard.edu/~hcs/

Dawson, T. L. (2002b). New tools, new insights: Kohlberg's moral reasoning stages revisited. *International Journal of Behavioral Development, 26,* 154–166.

Dawson, T. L. (2002). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement*. , 3, 146–189.

Dawson, T. L., Commons, M. L., & Wilson, M. (manuscript submitted June 2003) The shape of development. Mmanuscript submitted for publication.

Dawson, T. L., & Gabrielian, S. (2003). Developing conceptions of authority and contract across the life span: Two perspectives. *Developmental Review*.

Dawson, T. L., Xie, Y., & Wilson, M. (2003). Domain-general and domain-specific developmental assessments: Do they measure the same thing? *Cognitive Development*.

Demetriou, A., & Efklides, A. (1994). Structure, development, and dynamics of mind: A meta-Piagetian theory. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind, and reasoning: Structure and development. Advances in psychology* (pp. 75–109). Amsterdam, The Netherlands: Elsevier.

Faust, D., & Arbuthnot, J. (1978). Relationship between moral and Piagetian reasoning and the effectiveness of moral education. *Developmental Psychology, 14,* 435–436.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87,* 477–531.

Fischer, K. W., & Bidell, T. R. (1998). Dynamic development of psychological structures in action and thought. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (5th ed., pp. 467–561). New York: Wiley.

Fischer, K. W., & Bullock, D. H. (1981). Patterns of data: Sequence, synchrony, and constraint in cognitive development. In K. W. Fischer (Ed.), *Cognitive development* (pp. 1–22). San Francisco: Jossey-Bass.

Fischer, K. W., & Rose, S. P. (1999). Rulers, clocks, and nonlinear dynamics: Measurement and method in developmental research. In G. Savelsbergh, H. vanderMaas, & P. van Geert Eds.), *Nonlinear developmental processes* (pp. 197–212). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Fischer, K. W., & Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology, 36,* 613–648.

Gilligan, C. (1977). In a different voice: Women's conceptions of self and of morality. *Harvard Educational Review, 47,* 481–517.

Ginsburg, H. P., & Opper, S. (1988). *Piaget's theory of intellectual development* (3rd ed.).

Englewood Cliffs, NJ: Prentice-Hall.
Halford, G. S. (1999). The properties of representations used in higher cognitive processes: Developmental implications. In I. E. Sigel (Ed.), *Development of mental representation: Theories and applications.* (pp. 147–168). Mahwah, NJ: Erlbaum.
Keller, M., Eckensberger, L. H., & van Rosen, K. (1989). A critical note on the conception of preconventional morality: The case of Stage 2 in Kohlberg's theory. *International Journal of Behavioral Development, 12,* 57–69.
Kitchener, K. S., & King, P. M. (1990). The reflective judgment model: Ten years of research. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), *Adult development* (Vol. 2, pp. 62–78). New York: Praeger.
Kitchener, K. S., King, P. M., Wood, P. K., & Davison, M. L. (1989). Sequentiality and consistency in the development of reflective judgment: A six-year longitudinal study. *Journal of Applied Developmental Psychology, 10,* 73–95.
Kohlberg, L. (1969). Stage and sequence: The cognitive–developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.
Kohlberg, L. & Candee, D. (1984). Stage and sequence: The cognitive developmental approach to socialization. In *The psychology of moral development: The nature and validity of moral stages* (pp. 7–169). San Francisco: Jossey-Bass.
Kuhn, D. (1976). Short-term longitudinal evidence for the sequentiality of Kohlberg's early stages of moral judgment. *Developmental Psychology, 12,* 162–166.
Kuhn, D. (2000). Does memory development belong on an endangered topic list? *Child Development, 71,* 21–25.
Levine, C. G. (1979). The form–content distinction in moral developmental research. *Human Development, 22,* 225–234.
Miller, J. G. (1988). Bridging the content-structure dichotomy: Culture and the self. In B. Michael Harris (Ed.), *The cross-cultural challenge to social psychology. Cross-cultural research* (pp. 266–281). Newbury Park, CA: Sage.
Okonkwo, R. U. N. (1997). Moral development and culture in Kohlberg's theory: A Nigerian (Igbo) evidence. *IFE Psychologia: An International Journal, 5,* 117–128.
Overton, W. (1998). Developmental psychology: Philosophy, concepts, and methodology. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (5 ed., pp. 107–188). New York: Wiley.
Overton, W. F., Ward, S. L., Noveck, I. A., & Black, J. (1987). Form and content in the development of deductive reasoning. *Developmental Psychology, 23*.
Pascual-Leone, J., & Goodman, D. (1979). Intelligence and experience: A neoPiagetian approach. *Instructional Science, 8,* 301–367.
Piaget, J. (1965). *The moral judgment of the child* (M. Gabian, Trans.). New York: Free Press.
Piaget, J. (1977). The stages of intellectual development in childhood and adolescence. In H. E. Gruber & J. J. Vonéche (Eds.), *The essential Piaget* (pp. 814–819). New York: Basic Books.
Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development* (T. Brown & K. J. Thampy, Trans.). Chicago: University of Chicago Press.
Puka, B. (Ed.). (1994). *The great justice debate: Kohlberg criticism*. New York: Garland.
Roberts, R. J., Jr. (1981). Errors and the assessment of cognitive development. In K. W. Fischer (Ed.), *Cognitive development. New directions for child development* (Vol. 12, pp. 69–78). San Francisco: Jossey-Bass.
Simpson, E. L. (1973). Moral development research: A case study of scientific cultural bias. *Human Development, 17,* 81–106.

Snarey, J. R., Reimer, J., & Kohlberg, L. (1985). Development of social–moral reasoning among Kibbutz adolescents: A longitudinal cross-cultural study. *Developmental Psychology, 21,* 3–17.

Stack, C. (1990). Different voices, different visions: Race, gender, and moral reasoning. In R. Ginsberg & A. Tsing (Eds.), *The negotiation of gender in American society* (pp. 19–27). Boston: Beacon Press.

Stevens, D. A. (2000, June). *Dynamic systems: Continuities and discontinuities with Piaget's theory of equilibration.* Paper presented at the annual meeting of the Jean Piaget Society, Chicago.

van Geert, P. (2000). The dynamics of general developmental mechanisms: From Piaget and Vygotsky to dynamic systems models. *Current Directions in Psychological Science, 9,* 64–68.

Walker, L. J. (1980). Cognitive and perspective-taking prerequisites for moral development. *Child Development, 51,* 131–139.

Walker, L. J. (1982). The sequentiality of Kohlberg's stages of moral development. *Child Development, 53,* 1330–1336.

Walker, L. J. (1989). A longitudinal study of moral reasoning. *Child Development, 60,* 157–166.

Walker, L. J., Gustafson, P., & Hennig, K. H. (2001). The consolidation/transition model in moral reasoning development. *Developmental Psychology, 37,* 187–197.

Werner, H. (1948). *Comparative psychology of mental development*. Chicago: Follett.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.

Received December 19, 2002

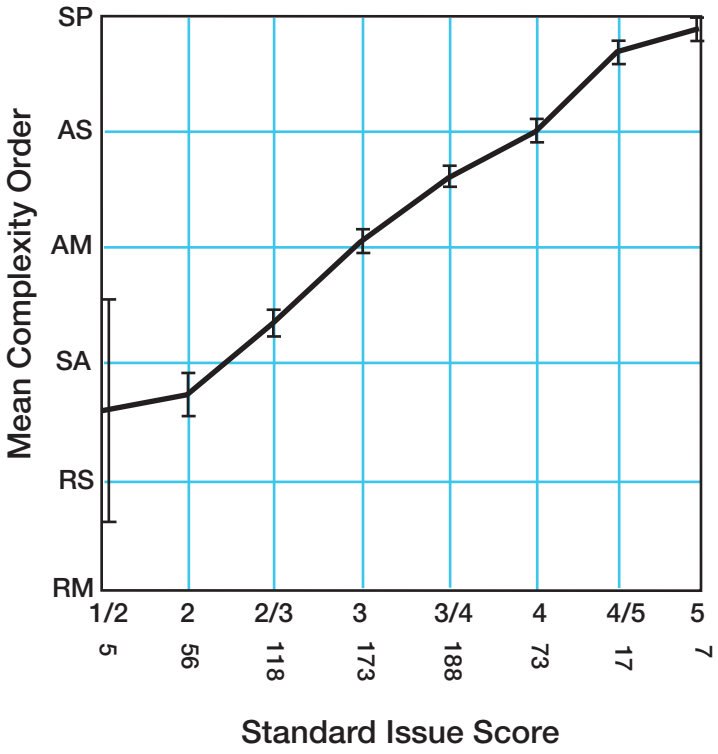

Figure 1: The relationship between Standard Issue and Hierarchical Complexity scores

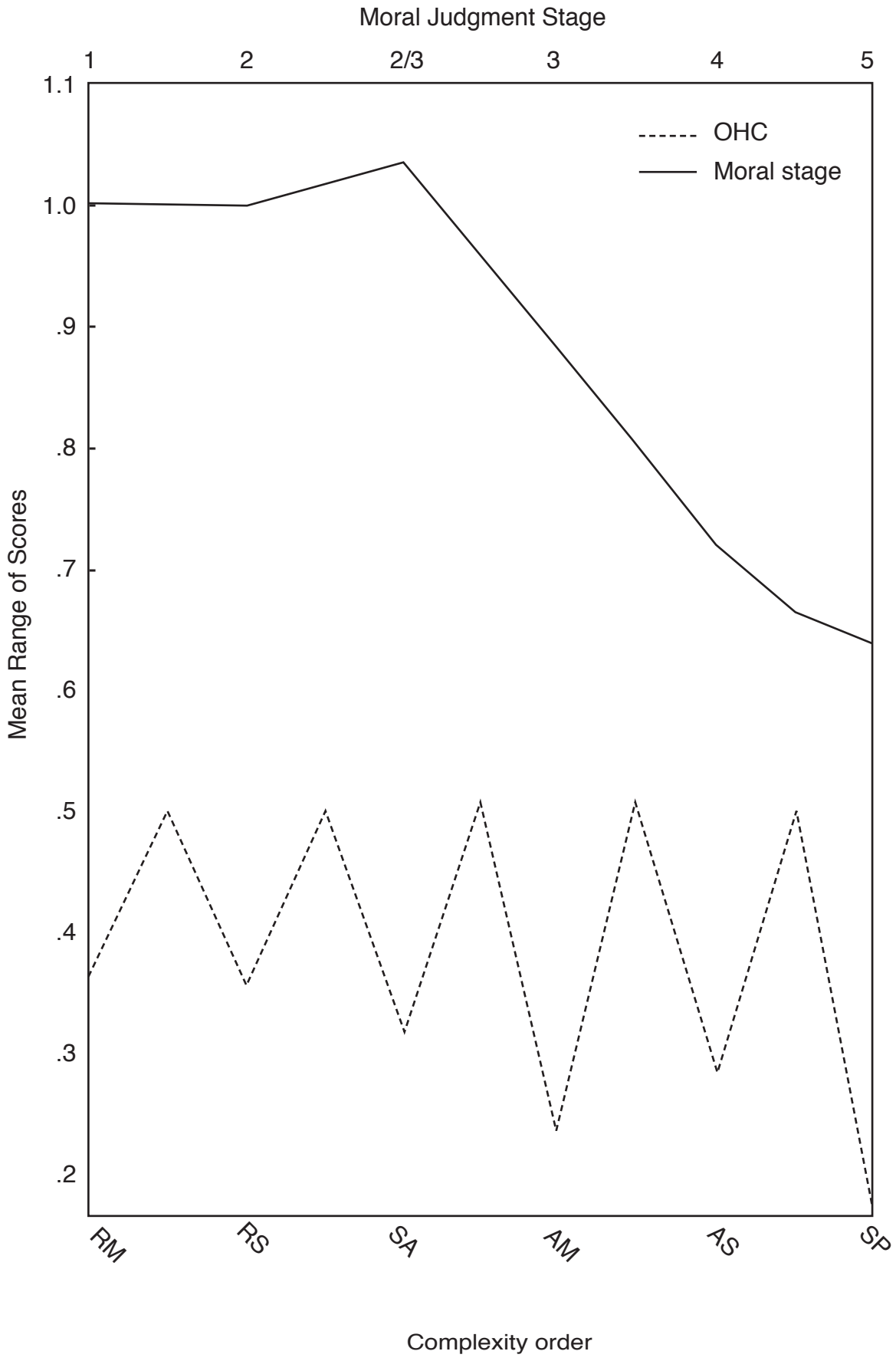

Figure 2: A comparison of mean score ranges at analogous moral judgment stages and complexity orders.

Table 2

Ages of attainment for Kohlbergian moral stages and complexity orders

Moral Stage							Complexity Order (CO)						
stage	n	mean	min	max	median	mode	**CO**	n	mean	min	max	median	mode
1.0	0						**RM**	36	5.6	5	7	5.5	5-6
1.5	5	9.6	7	14	10.0	7,10	**RM/RS**	20	6.2	5	7	6.0	6-7
2.0	56	12.1	6	44	10.0	10	**RS**	46	7.8	6	11	7.0	7
							RS/SA	38	9.9	7	15	10.0	10
2.5	118	17.3	7	54	14.0	16	**SA**	72	11.0	8	17	10.0	10
							SA/AM	54	16.4	8	55	14.0	14
3.0	173	27.1	10	68	19.0	17	**AM**	220	24.8	10	68	18.0	16
3.5	188	32.7	14	68	35.0	17	**AM/AS**	91	34.5	17	62	38.0	18
4.0	73	41.4	18	86	40.0		**AS**	151	39.5	17	82	40.0	22
4.5	17	38.2	21	69	38.0		**AS/SP**	22	41.7	20	86	39.9	
5.0	7	45.7	35	57	45.0		**SP**	27	41.0	21	64	41.0	

Table 3: Hierarchical complexity scores assigned to Criterion Judgments from Form A of the Standard Issue Scoring System

	Moral Stage								
	1	**1/2**	**2**	**2/3**	**3**	**3/4**	**4**	**4/5**	**5**
	13	8	25	17	40	26	46	24	20
Complexity Order									
RM	**6 (46%)**	1 (13%)							
RM/RS	4 (31%)	4 (50%)	2 (8%)						
RS	3 (23%)	3 (38%)	**10 (40%)**						
RS/SA			8 (32%)						
SA			5 (20%)	**16 (94%)**	2 (5%)				
SA/AM					5 (18%)				
AM				1 (6%)	**33 (83%)**	19 (73%)			
AM/AS						6 (23%)	1 (02%)		
AS						1 (4%)	**33 (72%)**	3 (13%)	
AS/SP							12 (26%)	2 (8%)	
SP								19 (79%)	**20(100%)**