DOCUMENT RESUME

ED 370 977 TM 021 550

AUTHOR Puncochar, Judith; And Others
TITLE When Two Heads Are Worse than One, Revisited: Confidence Resolutions by Individuals in Structured Learning Groups.
PUB DATE Apr 94
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Confidence Testing; *Decision Making; Graduate Students; Group Dynamics; Higher Education; Knowledge Level; Metacognition; Multiple Choice Tests; *Self Concept; Small Group Instruction; Testing Problems; Test Results; Undergraduate Students
IDENTIFIERS Accuracy; Majority Role; *Overconfidence

ABSTRACT

Individual and group assessments of quiz accuracy and students' discrimination of what they know and what they do not know regarding course material were examined using confidence ratings from 22 graduate students, 47 undergraduates, and their 23 heterogeneous learning groups over 6 quizzes. Students first took each multiple choice quiz as individuals and then as a group. Students received instruction regarding metamemory, confidence calibrations, and overconfidence after the first three quizzes. It was hypothesized that individuals and their groups would use this information to adjust their confidence ratings to discriminate appropriately between correct and wrong quiz answers. Within groups, students improved their accuracy, but did not appropriately adjust their confidence judgments. Moreover, the improved accuracy in groups came at a cost of increased confidence for wrong answers. Neither relevant information about metamemory nor assignment to structured learning groups was effective at improving students' assignments of confidence judgments, and may even have made it worse. Factors affecting group decision making appear to be high individual confidence and a majority effect, with educational status a marginally contributing component. There are six figures and two tables. (Contains 19 references.) (Author/SLD)

When Two Heads Are Worse Than One, Revisited: Confidence Resolutions by Individuals in Structured Learning Groups

Judith Punćochaŕ

Paul W. Fox

Mitchell Madson

University of Minnesota

American Educational Research Association

1994 Annual Meeting

Division J - Postsecondary Education

New Orleans, Louisiana

April 7, 1994

Running head: OVERCONFIDENCE IN SMALL GROUPS

Abstract

This investigation examines individual and group assessments of quiz accuracy and students' discrimination of what they know and what they do not know regarding course material studied over an academic quarter of instruction. Data include confidence ratings from graduate students (N = 22), undergraduates (N = 47), and their heterogeneous learning groups (N = 23) over a series of six quizzes. Students first took each 10-item multiple choice quiz as individuals, and then answered the same items as a group. Students received instruction regarding metamemory, confidence calibrations, and overconfidence after the first three quizzes. We hypothesized that individuals and their groups would use this information to adjust their confidence ratings to appropriately discriminate between correct and wrong quiz answers. Within groups, students improved their accuracy, but did not appropriately adjust their confidence judgments. Moreover, the benefit of improved accuracy in groups came at a cost of increased confidence for wrong answers! Neither relevant information about metamemory nor assignment to structured learning groups were effective at improving students' assignment of confidence judgments, and may even make it worse. Factors affecting group decision-making appear to be high individual confidence and a majority effect, with educational status as a marginally contributing component.

When Two Heads Are Worse Than One, Revisited: Confidence Resolutions by Individuals in Structured Learning Groups

Studies across disciplines have shown that people tend to be overconfident when assessing the accuracy of their knowledge (e.g., Lichtenstein & Fischhoff, 1977). A parallel tendency toward overconfidence in educational settings has been demonstrated to befall faculty (Fox & LeCount, 1991; Goranson, 1989), teaching assistants (Fox & LeCount, 1991; LeCount, Fox, & Beattie, 1993), postsecondary tutors (Beattie, 1992; LeCount, Fox, & Beattie, 1993), and postsecondary students (LeCount & Fox, 1993). Investigators have attempted to curb the overconfidence phenomena in a variety of ways. For example, listing reasons why an answer may be wrong before giving a confidence rating tends to improve the appropriateness of the confidence judgments (Koriat, Lichtenstein, & Fischhoff, 1980). However, other attempts to moderate overconfidence have met with limited success. Placing students in small structured learning groups, where differing answers and opinions should signal obvious uncertainty to group members, had no beneficial effect on reducing overconfidence in individuals and their groups (LeCount & Fox 1992). Worse yet, the resolution of confidence, defined operationally as the ability to discriminate between what one knows and what one does not know (by assigning significantly higher confidence for correct answers than for wrong answers), was actually poorer for students working in small structured learning groups than it was for individuals working alone (LeCount & Fox, 1992). We refer to this outcome, after LeCount and Fox (1992), as the "two heads" effect.

Thus, the acclaimed benefits of students working in groups (see, Sharan, 1980; Johnson & Johnson, 1989) appear to come at the cost of an unwanted *increase* in students' confidence that their group decisions are correct when, in fact, they are wrong!

The primary objectives of the present inquiry are threefold. First, we examine the largely untested hypothesis that exposure to scientific information in the form of readings and lecture pertaining to human metamemory in general and confidence calibration, confidence resolution, and overconfidence in particular will induce more appropriate confidence ratings. We predict that the inclusion of lectures and readings concerning theories of memory and metamemory (e.g., Leonesio & Thomas, 1990; Schacter, 1986; Zechmeister & Nyberg, 1982, chap. 11), confidence calibration, and the problem of overconfidence in individuals' assessments of what they do and do not know will improve confidence judgments and resolution for individuals and groups. Second, we seek to extend previous findings that the benefits of working in structured learning groups comes at a cost of increased confidence for wrong answers (LeCount & Fox, 199'). In line with these earlier findings, and regardless of the objective reality and potential value of differing answers in moderating confidence appropriately, we expect that students working in groups will increase their confidence for the wrong answers! Third, we seek to identify some of the probable group dynamics underlying group answer selection and the inappropriate assignment of confidence when groups are wrong. In particular, we examine the role of variables such as prior individual student accuracy, student confidence, educational status

(i.e., graduate and undergraduate standing) and gender as factors potentially contributing to those cases where groups are highly confident, but wrong.

The present research draws its theoretical perspectives from the fields of: (1) human metamemory and calibration of confidence, and (2) group dynamics and decision-making processes. These are briefly discussed in turn below.

Most people tend to be overconfident by judging that their answers to questions have a higher probability of being correct than is actually the case (e.g., Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, Lichtenstein, & Fischhoff, 1980). This tendency for unwarranted confidence is thought to be caused in part by processes of inference and reconstruction during memory retrieval that introduce error (Fischhoff, Slovic, & Lichtenstein, 1977). However, this tendency towards overconfidence can be troublesome because feelings of confidence in one's present knowledge may influence the pursuit of new information and interpretation of that information (Fischhoff, Slovic, & Lichtenstein, 1977), which makes the study of confidence in one's knowledge particularly relevant to educational settings.

Within our postsecondary educational system, many colleges and universities use structured cooperative learning groups in an effort to improve student learning (Cooper, et al., 1990; Johnson, Johnson, & Smith, 1991) and increase the accuracy of their learning (LeCount & Fox, 1992). Investigating the strengths, limitations, and dynamics of small groups is important for understanding their effective use in college classrooms. Cooperative groups provide a structure for

analyzing the reduction of uncertainty in small group social interactions.

LeCount and Fox (1992) examined the effects of structured cooperative learning groups on confidence calibrations. They hypothesized that the use of small learning groups within the classroom should help students not only to improve their accuracy but also adjust their confidence ratings to more effectively represent the true nature of their decisions. LeCount and Fox (1992) reasoned that a group comprised of individuals with differing opinions should signal some obvious uncertainty to the group members. Hence, groups with differing answers were expected to give lower confidence ratings on their quiz answers under such conditions. Surprisingly, they found a startling tendency toward an increase in confidence for the groups' wrong answers. Over a series of eleven quizzes, students working as individuals and in groups tended to become more (rather than less) confident when wrong! Indeed, group mean confidence when wrong often exceeded individual mean confidence when correct. Figure 1 compares the LeCount and Fox (1992) predicted (and idealized) results (panel a) with the obtained results for both individual and group confidence conditionalized on correct or wrong answers (panel b).

Insert Figure 1 about here.

The current research expands the previous inquiry by LeCount and Fox (1992) and examines the effects of course information about metamemory and confidence on the subsequent confidence ratings of

individuals and groups. We hypothesize that under conditions of exposure to metamemory information, both individuals and their groups should be able to assign higher confidence ratings when their answers are correct and lower confidence ratings when their answers are wrong. Specifically, we anticipate a broadening of the confidence resolution curves for both individuals and groups. That is, we expect confidence in correct answers to increase and confidence in wrong answers to decrease after exposure to relevant course readings and lecture information. Figure 2 illustrates the current study's predictions for confidence.

Insert Figure 2 about here.

Moreover, we investigate what effect, if any, the factors of educational status (i.e., graduate students or undergraduates), gender, and confidence have on accuracy and confidence decisions for individuals and groups. We examine the dynamics of group decision-making by comparing the individuals' prior answer selections and confidence ratings with the group answer selection and confidence rating. We expect that graduate students will influence group answer selection significantly more than undergraduates. We base this expectation on the assumption that graduate and undergraduate students alike will expect a broader range of knowledge and experience to exist among graduate students. Similarly, research on gender expectations, influence, and self-esteem in classroom settings (see Sadker, Sadker, & Long, 1993; Sadker & Sadker, 1994), suggests that male students are more influential than female

students. Therefore, we expect that male students will be more effective than female students in determining the group answer selection. Furthermore, given the situation where one or more group members are highly confident, it is expected that the highly confident individual(s) will influence group answer selection more often than less confident individuals. Examination of educational status, gender, and confidence will bring further insight into the dynamics of small group decision-making processes and provide additional knowledge for postsecondary educators about the effective use of learning groups in college classrooms.

With respect to accuracy, group members are expected to identify and correct individual errors, resulting in higher quiz mean accuracy scores for group work than for individuals working alone, irrespective of exposure to metamemory information. Thus, the present research examines: (1) the accuracy of individual and group decisions on actual course material, (2) the resolution of confidence by individuals and groups after exposure to informative readings and lecture materials about metamemory, and (3) the influence of certain individual characteristics on group decisions.

Method

The research participants were 69 students in an upper level undergraduate/graduate psychology course at the University of Minnesota (undergraduates: 30 women and 17 men; graduate students: 14 women and 8 men). Students received one quiz weekly from Week 3 through Week 8 during a ten-week term, for a total of six quizzes. The quizzes were given at the beginning of each class. Each quiz consisted of 10 four-alternative multiple choice items. Each item

was followed by a confidence judgment scale that marked the limits of confidence in terms of the expected percentage of being correct about the answer (see Adams, 1957). For example, the confidence scale for the four-alternative multiple choice items began at 25% (representing a completely chance level of confidence) and extended to 100% (indicating absolute certainty of correctness). The students were instructed to mark at chance levels if they were completely unsure of an answer and they were told not to mark below chance at any time. Students were assured that marking low confidence levels would *not* influence their score in the class.

The students took each quiz individually and marked their confidence scores on their individual answer sheets. After turning in their individual answer sheets, they took the same exam again in structured cooperative learning groups composed of three students. The instructor assigned students to heterogeneous groups (N = 23) based on gender and status (undergraduate v. graduate). Students remained with their assigned groups throughout the quarter. Within these groups, students were instructed to reach consensus about their group answers *and* their confidence judgments. All members of the group participated in the group quiz for bonus points to be added to their individual scores.

Before taking each quiz, the instructor gave students verbal and written instructions to mark their confidence scores as accurately as possible and to depress their confidence judgments if they felt the slightest bit uncertain in their answer. Feedback on the appropriateness of the students' selected answers was provided by the instructor in lecture format immediately after the group exam.

Readings on metamemory, overconfidence, and confidence calibrations were assigned before Quiz 4 (specifically, Leonesio & Nelson, 1990; Schacter, 1986; Zechmeister & Nyberg, 1982, chap. 11). A lecture on metamemory, confidence calibrations, and overconfidence preceded Quiz 5. Thus, Quizzes 1, 2, and 3 were taken by students before exposure to information about confidence judgements, and Quizzes 4, 5, and 6 reflect involvement with relevant information about confidence resolution and calibration.

Results

Our investigation affirmed some of our predictions, but yielded several surprises. We expected that accuracy would improve when students worked in their learning groups, and it did. We also expected, based on findings by LeCount and Fox (1992), that the "Two Heads" phenomenon would begin to emerge during the first three quizzes, such that students working in structured learning groups would become more confident, not less, in their wrong answers, compared to individuals' confidence levels. In addition, we expected that the spread between group confidence when correct and confidence when wrong would shrink. This trend was observed and serves as a partial replication of the LeCount and Fox (1992) "two heads" effect. However, we also expected that the class readings and lecture on metamemory, confidence calibration, and confidence resolution would counter the "Two Heads" effect over Quizzes 4-6, and this did *not* happen. Instead, inexorably, group confidence when wrong kept rising over the quizzes, even to the point of exceeding individuals' confidence when correct! These results are described in turn below.

Accuracy

The six quiz means were relatively high across exams for both the individuals and their groups ($\bar{X}_{(ind)} = .80$; $\bar{X}_{(gr)} = .93$). We expected that the quiz accuracy means would be consistently higher for structured learning groups than for students working as individuals, and they were ($t_{(5)} = 14.05$, $p < .01$). As shown in Figure 3, group mean accuracy scores were higher than individual mean accuracy scores on every quiz (see LeCount & Fox, 1992, for similar findings).

Insert Figure 3 about here.

Moreover, Figure 3 shows highly consistent mean item difficulties over the six quizzes. Mean accuracy scores were essentially homogeneous for groups on all quizzes. Likewise, individuals received virtually equivalent mean accuracy scores, with the exception of quiz six which was slightly more difficult.

Graduate students scored consistently higher than undergraduates on all quizzes ($\bar{X}_{(Grad)} = .84$; $\bar{X}_{(UG)} = .77$), but these higher scores were only marginally significance ($F_{(1,65)} = 3.821$, $p = .055$). There were no gender differences or interactions between gender and educational status for accuracy ($F_{(1,65)} < 1$).

Overall Confidence

The findings on overall mean confidence, irrespective of whether the individual or group answer was right or wrong, tended to mirror the findings on overall accuracy. Mean individual confidence over

the six quizzes was 71.8%, which was significantly lower than the overall mean confidence of 87.8% for the same students working in their groups ($t_{(5)} = 20.66$, $p < .01$). Men and women gave similar overall individual confidence ratings (men = 74%; women = 71%) and showed no gender differences ($F_{(1,65)} < 1$). Moreover, the difference between overall assignment of confidence by the graduates ($\bar{X}_{(Grad)}$ = 76%) and undergraduates ($\bar{X}_{(UG)}$ = 70%) approached but did *not* reach significance ($F_{(1,65)} = 3.297$, $p = .074$), which is consistent with the undergraduates' marginally significant lower accuracy scores.

Confidence Judgments Correlated with Accuracy

Both individuals and their groups showed an increase in confidence which corresponded with an increase in accuracy. That is, higher mean confidence ratings were associated with higher mean accuracy levels ($r_{(ind)} = .72$, $p < .01$; $r_{(gr)} = .53$, $p < .01$). Likewise, individuals' correlations of accuracy scores with confidence ratings before and after exposure to course readings on metamemory ($Time_1$ and $Time_2$, respectfully) were similar and relatively high ($r_{(Time1)} = .68$, $p < .01$; $r_{(Time2)} = .66$, $p < .01$). However, a different picture emerges for the groups. Before exposure to the relevant course readings, groups demonstrated an association between confidence and accuracy that was comparable in magnitude to that of the individuals (groups: $r_{(Time1)} = .66$, $p < .01$). However, the groups at $Time_2$ demonstrated a very poor confidence/accuracy association (groups: $r_{(Time2)} = .24$, ns). The groups' inability to assign appropriate confidence judgments during $Time_2$ is startling for two reasons: (1) the group quiz difficulty levels during $Time_1$ and

$Time_2$ were similar, and (2) $Time_2$ confidence judgments came *after* exposure to informative readings and lecture materials.

Confidence Judgments Conditionalized on Correct or Wrong Answers

Confidence judgments were classified in terms of those associated with correct answers and those associated with wrong answers. As Table 1 shows, students working as individuals or in groups demonstrated overall resolution of confidence, in that they gave confidence ratings that were higher when their answers were correct than when their answers were wrong ($t_{(68)} = 16.47$, $p < .01$; $t_{(21)} = 7.40$, $p < .01$). Similarly, both before and after exposure to information about confidence calibration, resolution, and overconfidence, both individuals and groups maintained good resolution of confidence ($Time_1$: $t_{(63)} = 16.73$, $p < .01$; $t_{(16)} = 4.78$, $p < .01$; $Time_2$: $t_{(66)} = 11.45$, $p < .01$; $t_{(18)} = 4.40$, $p < .01$).

Insert Table 1 about here.

Group mean confidence when correct exceeded individuals' confidence when correct on every quiz ($t_{(5)} = 22.637$, $p < .01$). However, with the exception of a slight increase in mean confidence for correct answers on the fourth quiz, Table 2 describes a gradual, but not significant, *decrease* in individual and group mean confidence when correct over the six quizzes.

Insert Table 2 about here.

Group mean confidence when wrong also exceeded individual mean confidence when wrong on every quiz ($t_{(5)} = 4.672$, $p < .01$). Indeed, the present study replicates and extends the earlier surprising findings of (LeCount & Fox, 1992), namely that groups' confidence when *wrong* continues to increase over the quizzes, even to the point of overtaking individuals' confidence when correct! Exposure to information in the form of readings and lecture about confidence calibrations and overconfidence may have a small moderating effect on the confidence judgments by individuals when they are wrong, but not on their groups' confidence judgments when their groups are wrong! Furthermore, group mean confidence when wrong reached its highest level *after* exposure to information on overconfidence! As Figure 4 demonstrates, participants in groups were more confident than participants working as individuals, especially when they are wrong!

Insert Figure 4 about here.

There were no gender differences in mean confidence when correct or when wrong ($t_{(67)} < 1$), but graduate students gave marginally higher confidence ratings when correct than undergraduates ($t_{(67)} = 1.98$, $p = .052$).[1]

Use of the Confidence Scale

---

[1]Both women and men showed good resolution of confidence judgments for correct and wrong answers ($t_{(44)} = 13.29$, $p < .01$; $t_{(23)} = 10.10$, $p < .01$). Likewise, graduate students and undergraduates demonstrated appropriate resolution of confidence judgments when correct and wrong ($t_{(22)} = 9.86$, $p < .01$; $t_{(45)} = 13.91$, $p < .01$).

We examined whether groups and individuals used the confidence scale dissimilarly. Figure 5 shows the groups' and individuals' differential use of the confidence scale. Individuals are more likely to spread their confidence judgments out evenly across the confidence scale. However, most of the groups' confidence judgments are clustered between 90% and 100%!

Insert Figure 5 about here.

Group Dynamics Underlying Answer Selection and Confidence Ratings

We examined a number of candidates for factors that influence group answer selection: 1) individual accuracy, 2) individual confidence, 3) a majority effect of similar (correct or wrong) answers, and 4) individual differences (i.e., gender and educational status). A total of 1155 group decisions of interest (both where groups are correct and groups are wrong) were examined to identify the singular and/or aggregate influencing factors within specific group profiles that may contribute to incorrect answer selection and inappropriate overconfidence in groups' wrong answers. First, we matched individuals' answers with the groups' answers to identify the group members who had the greatest "voice" in determining the group decision. Individuals whose answers matched the groups' decision received a score of +1. Unmatched answers received a score of zero. Mean match scores for gender showed that there were no differences in matches between women and men with their groups' decisions ($F_{(1,65)} = 1.376$, $p = .25$). Graduates students approached a greater number of matches than undergraduates, but these matches did

not reach statistical significance ($F_{(1,65)} = 3.171$, $p = .08$). There were no interactions between gender and educational status.

We examined various group profiles defined by group members' personal answer selections (e.g., when three individuals in a group were all correct (CCC), correct-correct-wrong, respectively (CCW); correct-wrong-wrong with the *same* wrong answers (CWW_{sim}); correct-wrong-wrong with dissimilar wrong answers (CWW_{dis}), etc.). Figure 6 shows that regardless of whether the group decision is correct or wrong, confidence in the group decision is highest when group members are consensual. The lowest level of confidence in the group decision, regardless of whether the decision is correct or wrong, occurs when all group members initially endorse differing answers. In between, the confidence of the groups when correct and when wrong is nicely ordered by a majority of the members having been correct or wrong on their prior individual answers.

Insert Figure 6 about here.

Extending the group profile investigation further, a leading factor that seems to influence whether the groups make correct or wrong decisions appears to be a majority effect. For example, in a group with a prior individual answer profile of CCW, the percentage of items on which groups answer correctly is 99%! The majority effect of having two correct answers seems to be a dominant influencing factor. However, the majority effect seems *not* to prevail in CWW_{sim} groups with similar wrong answers. CWW_{sim} groups chose a correct answer 52% of the time. The factor that appears to

lessen the majority effect of same wrong answers appear to be the confidence of the individual with the correct answer. Figure 7 shows that the group usually chooses correctly when the individual with a correct answer has higher confidence than the mean confidence of the group members with matching wrong answers ($\bar{X}_{(ind.correct)} = 80.3$; $\bar{X}_{(inds.wrong)} = 57.8$; $t_{(28)} = 6.75$, $p < .01$). Conversely, when the individuals with a correct answer has low confidence, the group usually opts for the majority's choice, albeit incorrect. The CWW_{sim} groups that select a wrong group answer show *no* significant difference in the mean confidence levels between individuals with correct or wrong answers ($\bar{X}_{(ind.correct)} = 58.5$; $\bar{X}_{(inds.wrong)} = 55.5$; $t_{(27)} < 1$). Groups opt for the correct decision when individuals with correct answers have a high mean confidence, rather than a moderate or low mean confidence ($t_{(55)} = 3.972$, $p < .01$). Likewise, the CWW_{dis} groups show a similar pattern of results.

Insert Figure 7 about here.

We extended the investigation of group profiles to a case by case approach in an effort to identify probable characteristics of groups that may contribute to the inappropriate assignment of high confidence when groups are wrong. The influencing factors of importance that contribute to high confidence when groups are wrong appear to be the individuals' ratings of confidence and a majority effect, with educational status as a marginally contributing component.

Concluding Discussion

This study combines the processes and research literatures of human judgment and metamemory (e.g., Lichtenstein & Fischhoff, 1977) with cooperative learning groups (e.g., Johnson, & Johnson, 1989) to replicate and extend previous research findings (LeCount & Fox, 1992). The specific instructional intervention of information about metamemory, confidence calibrations, and overconfidence appears to dampen individuals' levels of confidence, but not that of groups. The basic finding suggests caution about a potentially undesirable by-product of people working in cooperative groups. Once again, the benefits of students working in groups (i.e., increased accuracy of the students' answers) appears to come at the cost of an unwanted increase in groups' confidence in their wrong answers. The educational significance of this research lies in what it tells the college instructor about the difficulty of helping students to more appropriately discriminate between what they understand and what they do not, especially in cooperative groups, irrespective of the other benefits of cooperative group work.

In summary, even when information is provided to students and their groups about calibration of confidence and the problems of overconfidence, confidence increases when groups are wrong. Moreover, this study replicates the previous findings (LeCount & Fox, 1992) that over time, small learning groups tend to become more and more convinced of their own understanding of the information even when they are wrong!

References

Adams, J. K. (1957). A confidence scale defined in terms of expected percentages. *American Journal of Psychology, 79,* 432-436.

Beattie, S. (1992). *Testing the misestimation hypothesis: How well do university instructors know what their students have learned?* Unpublished honour's thesis, York University, North York, Ontario.

Cooper, J., Prescott, S., Cook, L., Smith, L., Mueck, R., & Cuseo, J. (1990). Cooperative learning and college instruction. Long Beach, CA: California State University Institute for Teaching and Learning.

Fox, P. W., & LeCount, J. (1991, April). *When more is less: Faculty misestimation of student learning*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552-564.

Goranson, R. E. (1989). *Are teachers well calibrated?* Paper presented at the annual meeting of the Canadian Psychological Association, Halifax, NS.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Active learning: Cooperation in the college classroom*. Edina, MN: Interaction Book Co.

Johnson, D. W., & Johnson, R. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book Company.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118.

LeCount, J., & Fox, P. W. (1992, April). *When two heads are worse than one: The Effect of Feedback and Structured Learning Groups on Confidence Calibrations*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

LeCount, J., Fox, P. W., Beattie, S. (1993, April). *When more is less, revisited: Reducing the phenomenon of instructor misestimation of student learning*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GE.

Leonesio, R. J., & Thomas, O. N. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory & Cognition, 16*, 464-470.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, (1990). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge, MA: Cambridge University Press.

Sadker, M., Sadker, D., & Long, L. (1993). Gender and educational equity. In J. A. Banks, & G. A. M. Banks (Eds.), *Multicultural education: Issues and perspectives* (2nd ed.) (pp. 111-127). Boston: Allyn & Bacon.

Sadker, M., & Sadker, D. (1994). *Failing at fairness: How America's schools cheat girls.* Toronto: Macmillan.

Schacter, D. L. (1986). Feeling-of-knowing ratings distinguish between genuine and simulated forgetting. *Journal of Experimental Psychology: Learning, Memory & Cognition, 12,* 30-41.

Sharan, S. (1980). Cooperative learning in teams: Recent methods and effects on achievement, attitudes, and ethnic relations. *Review of Educational Research, 50,* 241-272.

Zechmeister, E. B., & Nyberg, S. E. (1982). *Human memory.* Monterey, CA: Brooks/Cole Publishing, 229-231; 239-249.

Table 1

Mean Confidence of Individuals and Groups Before and After Exposure to Metamemory Information

Confidence	Correct	Wrong
Individual Overall	75.0* (13.9)	56.8* (14.1)
$Time_1$ (Quizzes 1-3)	76.0 (14.4)	58.9 (15.1)
$Time_2$ (Quizzes 4-6)	73.5 (14.7)	55.5 (15.1)
Group Confidence	89.2* (6.0)	67.8* (14.0)
$Time_1$ (Quizzes 1-3)	90.5 (5.7)	68.3 (19.2)
$Time_2$ (Quizzes 4-6)	88.3 (8.2)	70.2 (18.5)

Note. Asterisks denote means that differ significantly at $p \leq .01$. Comparisons between confidence when correct and confidence when wrong at $Time_1$ and $Time_2$ are significant at $p \leq .01$.

Table 2

Mean Confidence of Students Working as Individuals or in their Groups When Correct and Wrong over the Six Quizzes

	Individuals		Groups	
	Cf.Correct	Cf.Wrong	Cf.Correct	Cf.Wrong
Quiz				
1	78.8 (21.3)	55.6 (21.1)	90.0 (8.3)	58.2 (21.1)
2	78.9 (20.3)	53.0 (19.5)	92.2 (6.4)	63.3 (18.6)
3	75.3 (21.8)	55.0 (19.7)	89.4 (8.1)	69.0 (18.9)
4	80.1 (19.7)	56.0 (21.6)	92.2 (6.4)	68.3 (14.4)
5	74.1 (22.2)	50.4 (20.8)	85.7 (16.4)	72.1 (22.0)
6	70.7 (21.9)	51.4 (21.0)	85.4 (14.5)	72.8 (20.7)
Overall	76.8* (21.3)	53.8* (20.8)	89.6* (15.5)	67.3* (20.8)

Note. Asterisks denote overall confidence means that differ significantly at $p \leq .05$.

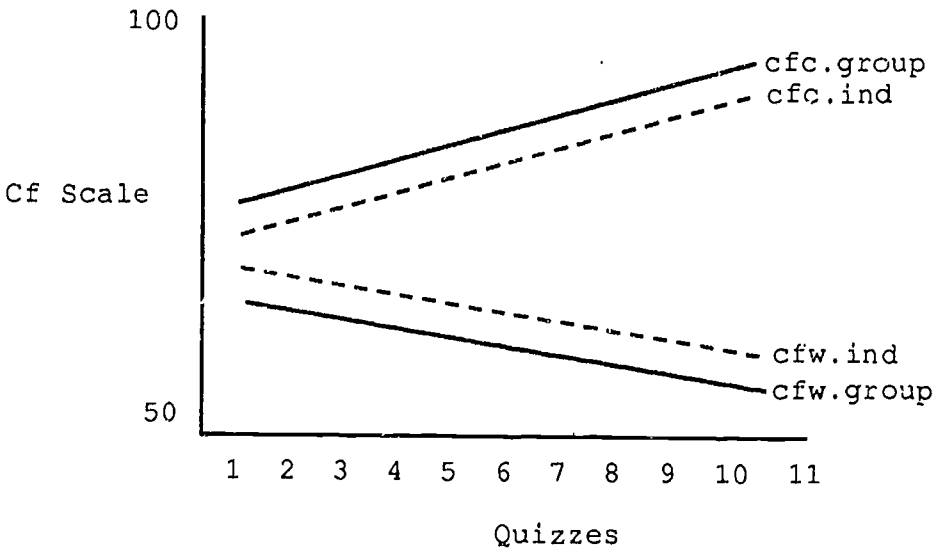


Figure 1a. Predicted individual and group confidence when correct (cfc) and when wrong (cfw).

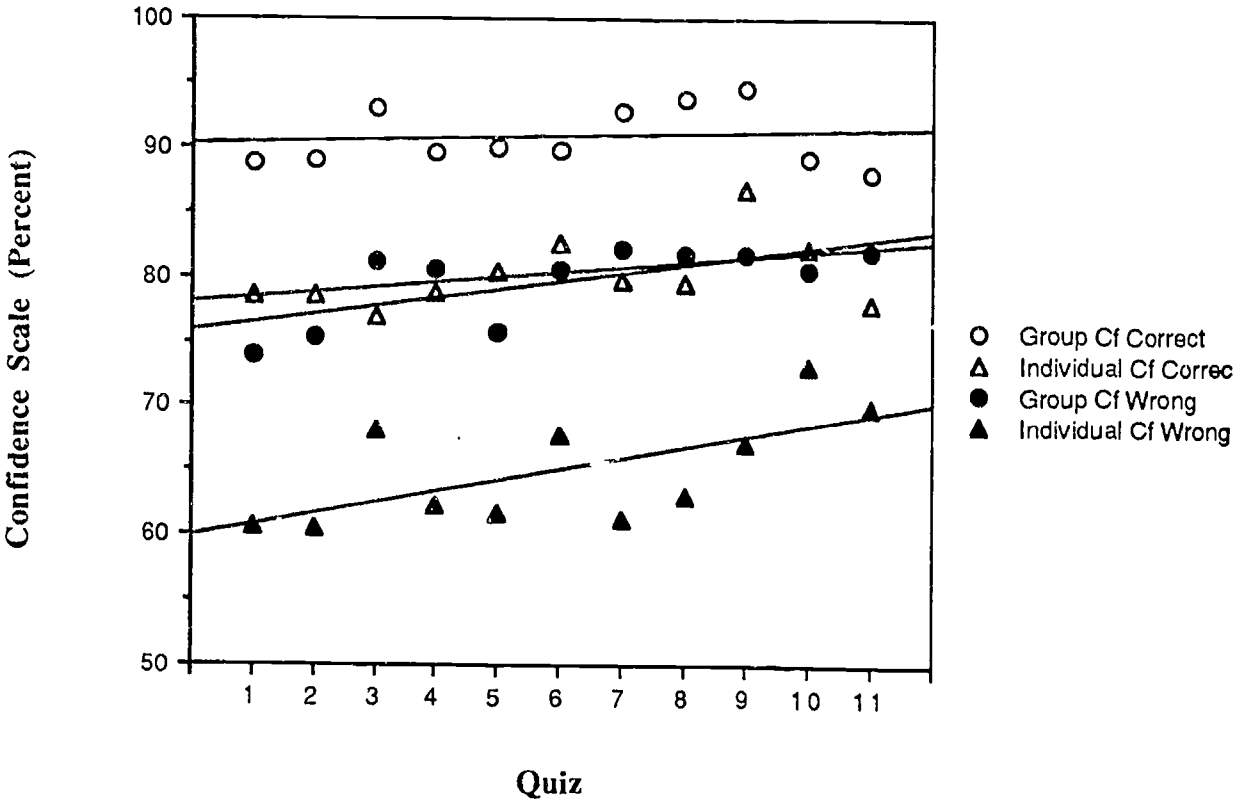


Figure 1b. Obtained individual and group confidence when correct (cfc) and when wrong (cfw).

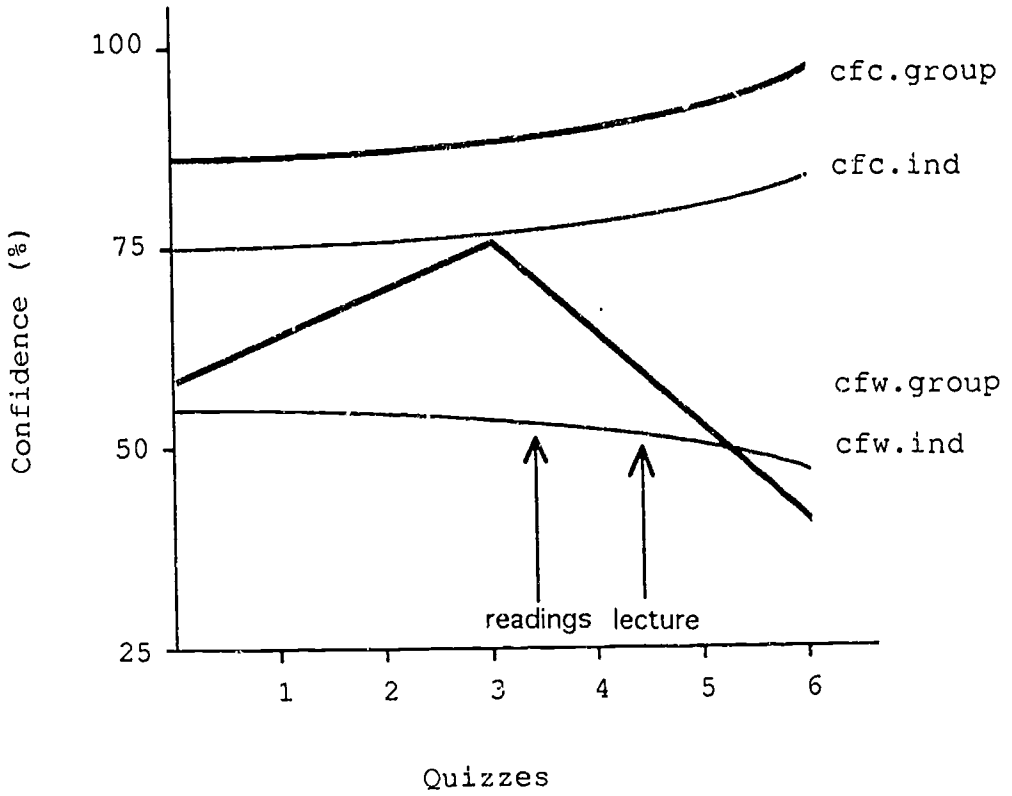


Figure 2. Predicted individual and group confidence when correct (cfc) and when wrong (cfw).

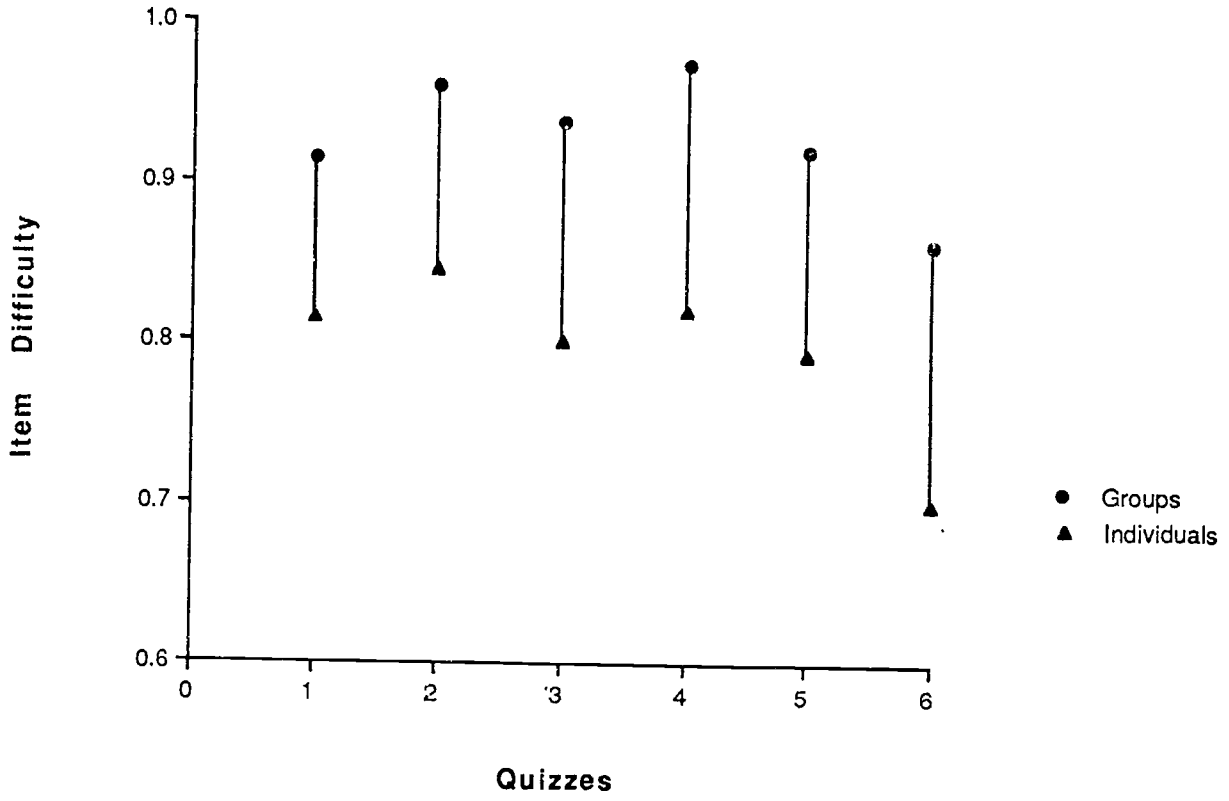


Figure 3. Mean item difficulty for individuals and their groups over the six quizzes.

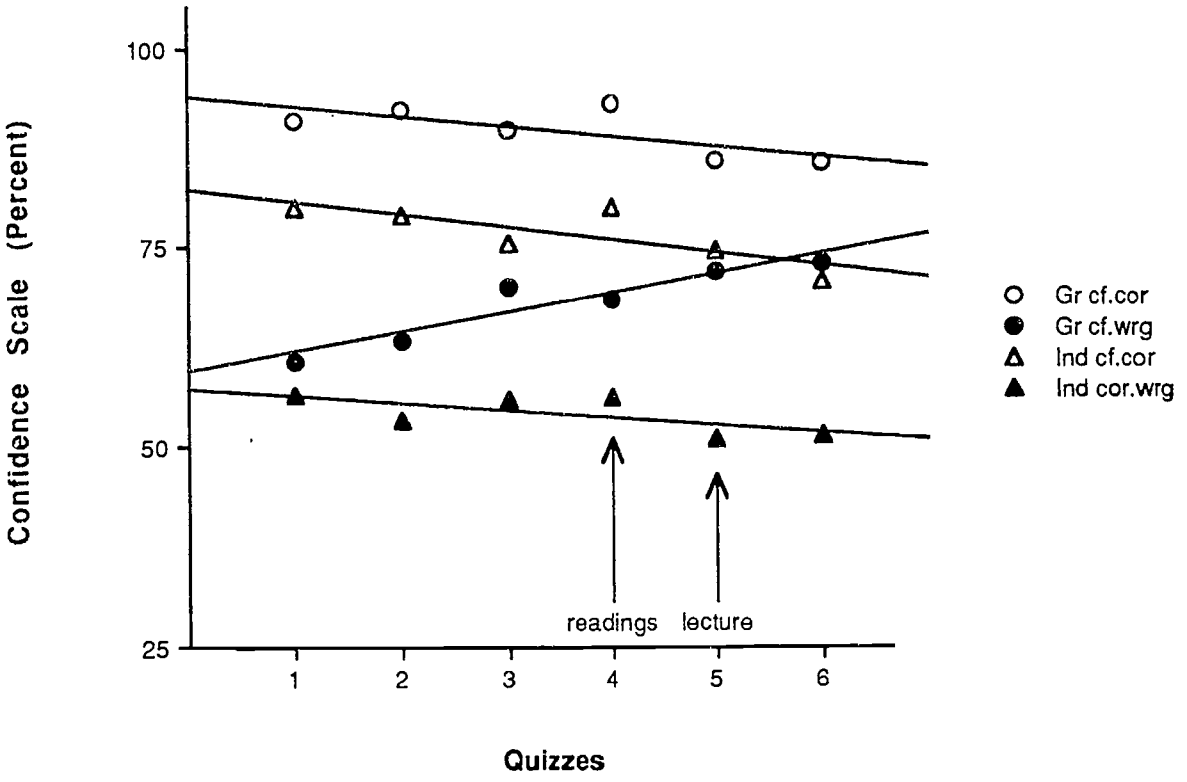


Figure 4. Mean confidence ratings when correct and wrong for individuals and their groups over the six quizzes.

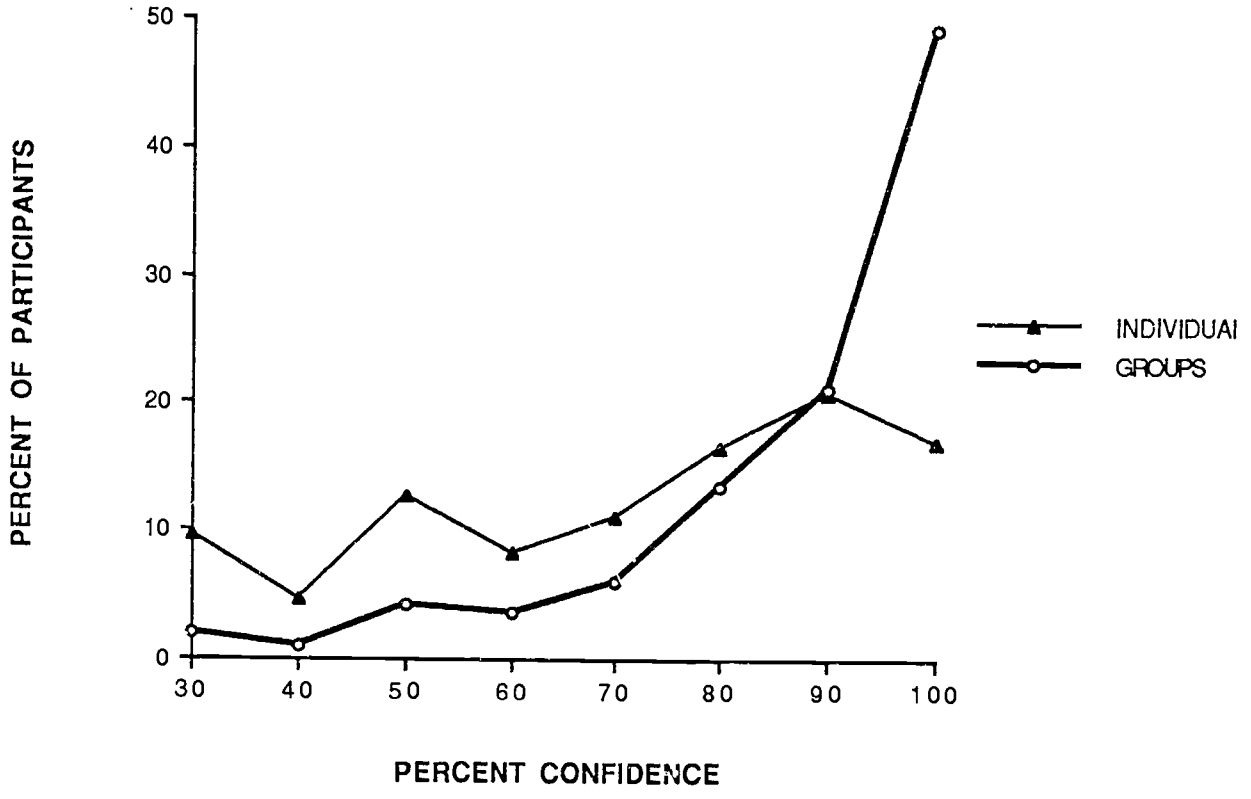


Figure 5. Use of the confidence scale by individuals and their groups.

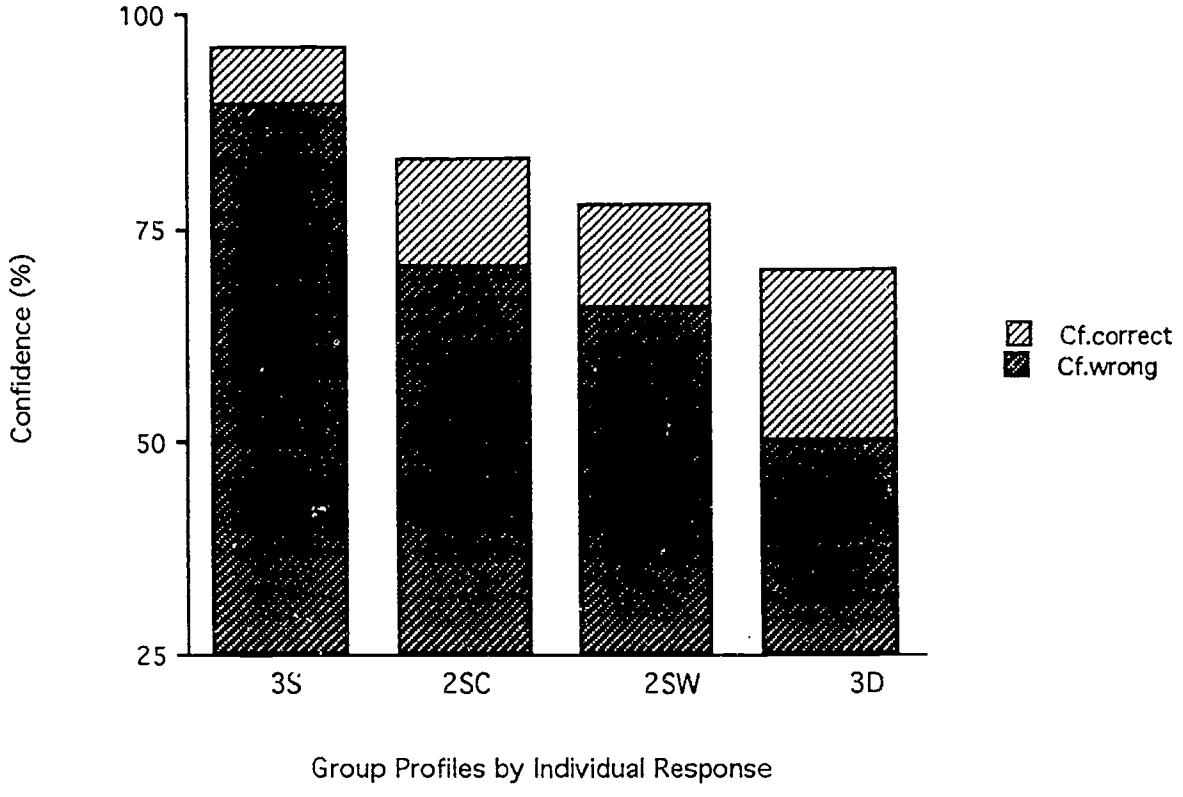
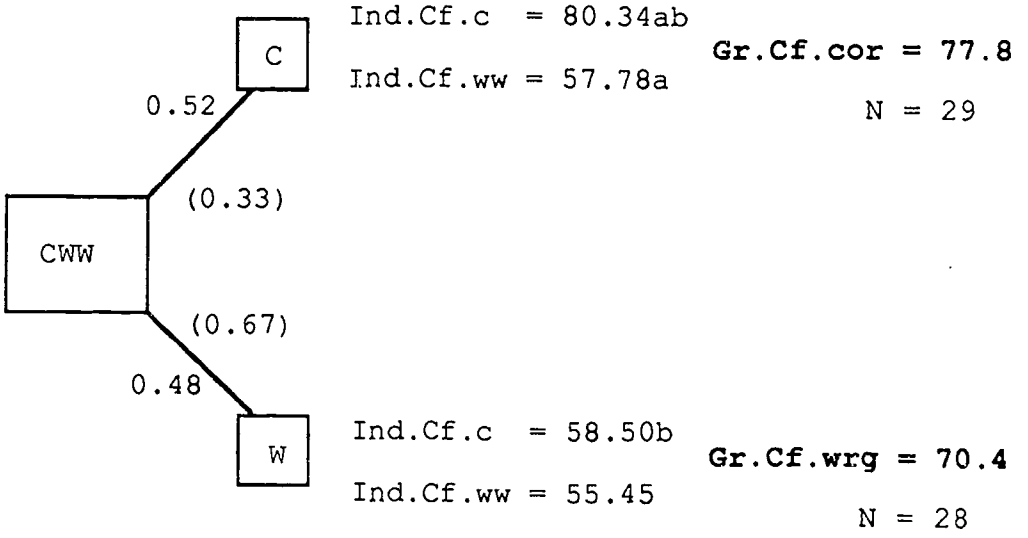


Figure 6. Group mean confidence ratings on quiz items by individual answer selection (i.e., 3S = three similar answers, 2SC = two similar correct answers, 2SW = two similar wrong answers, 3D = three differing answers).

Similar Wrong Answers



Dissimilar Wrong Answers

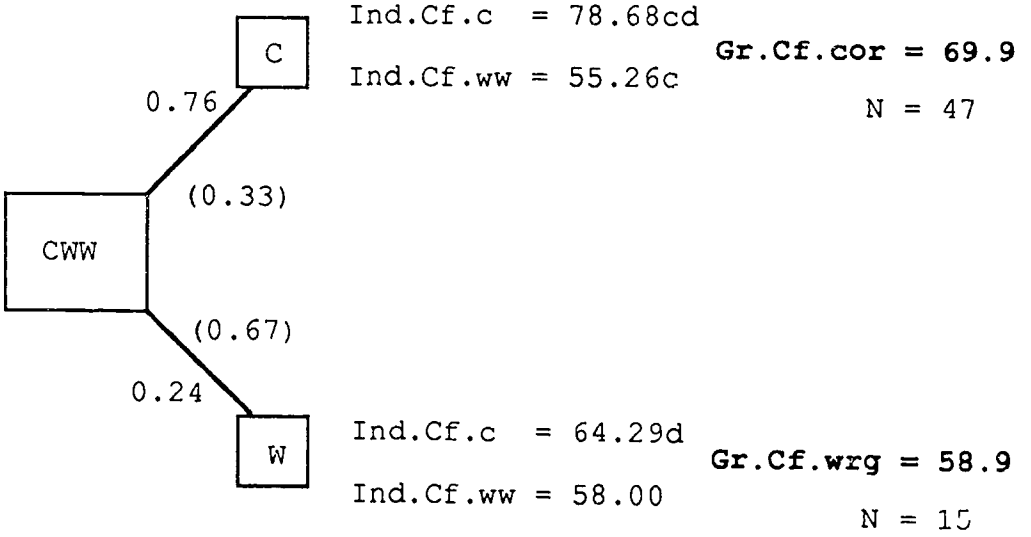


Figure 7. Accuracy probabilities and mean confidence scores for individuals and their groups. Expected probabilities are in parentheses. Means with matching subscripts differ significantly at $p \leq .05$.