

Gender gaps in group listening and speaking: issues in social constructivist approaches to teaching and learning

Darryl Hunter^{*a}, Trevor Gambell^b and Bikkar Randhawa^b

^a*University of British Columbia, Canada;* ^b*University of Saskatchewan, Canada*

Because of its centrality to school success, social status, and workplace effectiveness, oral and aural skills development has been increasingly emphasized in Canadian curricula, classrooms and, very recently, large-scale assessment. The corresponding emphasis on group processes and collaborative learning has aimed to address equity issues in schools. However, a 1998 Canadian assessment of students' speech communication skills ($N = 551$ groups) yielded many significant gender differences in individual listening skills, group production, and self-efficacy. The oral production in small groups of majority- or all-male groups lagged significantly behind that of all-female groups. The girl–boy gaps in oracy parallel those evident for literacy in provincial (state), national and international studies among adolescents. Implications are drawn for social constructivist pedagogy, for curricular, instructional, and evaluation practices, and for redressing gender differences.

Introduction

Our purposes in this article are to explore gender questions in small group learning situations, and to consider issues in oracy skills development within a social constructivist framework. Our objectives are three-fold: to investigate girl–boy gaps in individual and group performances in listening and speaking from a Canadian large-scale, language arts (English)¹ assessment; to consider variables within groups of various male–female composition engaged in oral language processes (as contrasted with products); and to determine whether there are gender differences in perceived self-efficacy about speech communication. Observed gender differences in group production and perceptions, we posit, may explain engendered outcomes from provincial, national and international assessments in language arts. Specifically, we pose four research questions:

*Corresponding author. Faculty of Educational Studies, University of British Columbia, Vancouver, Canada. Email: darrylinvic@hotmail.com

- (1) To what extent are there gender differences in individual students' listening skills?
- (2) How does gender group composition interrelate with the quality of discussion in a social context?
- (3) To what extent are there girl-boy gaps in students' perceived self-efficacy in speech communication?
- (4) What are the gender patterns in students' self- and peer-evaluations of group performance after engaging in small group discussion?

Language arts specialists have become interested in listening and speaking abilities within group contexts because of changes in curricular assumptions during the past two decades in Canada. Curriculum writers have moved language arts pedagogy from behavioural and cognitive foundations with an almost exclusive focus on literacy (reading and writing) toward a social constructivist approach that encompasses all communicative modes (Governments of Alberta, British Columbia, Manitoba, North-west Territories and Yukon Territory, 1998). Although the ideas of Dewey, Vygotsky, and Bruner underpin much of this shift, constructivism in English language arts education might be said to have originated in the Bullock Report, *A Language for Life* (DES, 1975) in England, and transmogrified in North America as language and learning across the curriculum. Hynds (1997) describes contemporary literacy education as social constructivist because of its sociological epistemology, which attends to the way knowledge is constructed by, for, and between members of a discourse community such as a classroom. Hynds also points to sociopolitical constructivism, grounded largely in critical theory. In this paper, we prefer the term social constructivism, avoiding the political manifestations of literacy and language in classroom situations, while affirming that critical literacy is still a concern.

Constructivists believe the term 'text' extends well beyond the written word. Oral texts and visual representations hold the potential for meaning that can, and most often is, conveyed linguistically. The teacher recognizes that all linguistic and paralinguistic signs require interpretation because they are acts and constructions integral to learning. A central precept in constructivist classrooms is that a student is both author and architect of knowledge, rather than being a passive receptacle for the knowledge of others (Cook-Sather, 2002). Constructivists support reader-response pedagogy and process orientations. Classrooms are activity-based, depend on student-centred methods, and rely on students' life experiences to inform their interpretation of texts. Cooperative group discussions and debates, team inquiry, and collaborative project work are recommended in curriculum guides. Small group processes such as jigsaw activities foster a wider band of communication skills by eliciting listening and speaking and non-verbal abilities. These instructional practices are touted as means for creating a more inclusive and egalitarian classroom, thereby supporting equality of opportunity (Barnes & Todd, 1977).

Gender, culture, and hegemony will play out in constructivist classrooms, we acknowledge, as within any learning environment. Baxter (1999) notes that 'essentialists' viewed gender as constructed within the singular individual and

expressed linguistically as a set of features or a style geared towards particular speech functions. Essentialist approaches created a tendency to generalize and dichotomize male and female speech styles, or the dominance of male over female talk (Aries, 1976, 1977). In contrast, Baxter proposes 'that gender be investigated as a social construction realized through social interaction and organized by power relations' (Baxter, 1999, p. 82). Accordingly, in this paper, we use the term 'gender' to describe the self-identified, physiological groups of male and female (i.e. 'gender' is used in place of 'sex') to avoid any essentialist misconceptions. We presume that speakers constantly need to renegotiate their subject identities within multiple and competing discursive practices. Investigating language and gender in this way entails the assessment of speech practices in different classroom contexts that may position both girls and boys as relatively self-efficacious or powerless.

Self-efficacy, as a key concept in recent social learning theory, has been defined as the belief in one's ability to execute successfully a certain course of behaviour, or to manage prospective situations (Bandura, 1995; Lee & Bobko, 1994). Self-referential thinking is critical to psychosocial functioning in four ways. First, it affects thought patterns. Because much of human behaviour is regulated by forethought, goal-setting is directly influenced by individuals' self-appraisal of their capacities; self-efficacy influences the type of anticipatory scenarios individuals will construct, allows individuals to remain task-oriented in the face of institutional demands, and enables individuals to predict how they can control events in their lives (Bandura, 1997). Second, self-efficacy beliefs play a key role in students' self-regulation of motivation by shaping their conception of ability, the perceived importance and appeal of the task at hand, and their anticipation of likely outcomes (Bandura, 1993). Third, self-efficacy beliefs influence individual's feelings, likes and dislikes, and coping capabilities with stress (Bandura, 1995). And fourth, self-efficacy will influence decisions about the way that individuals choose to engage or avoid environmental constraints (Bandura, 1995). Academic self-efficacy is 'defined as personal judgments of one's capabilities to organize and execute courses of action to attain designated types of educational performances' (Zimmerman, 1995, p. 203). Self-efficacy will influence the choice of whether to engage in a task, the effort and persistence demonstrated in accomplishing it, and the standard of performance in appraising it (Bandura, 1993). Self-efficacy is thus interrelated with the learner's success.

Perhaps ironically, a number of provincial (state), national, and international studies during the last decade have drawn into question the presumption of gender equitable classrooms as both curricula and assessments incorporate a broader range of language modes and processes. For example, the Organization of Economic Cooperation and Development's (OECD, 2002) study of reading in the year 2000 across 32 countries proceeded from explicitly constructivist premises. It revealed female superiority in all participating countries, including the US, Canada and the UK, in all dimensions of reading performance. Fifteen-year-old girls demonstrated greater persistence in finishing reading tasks, more positive attitudes toward reading, more time spent reading for enjoyment, and more effective learning strategies, such

as adopting self-evaluative perspectives and elaborated planning, organizing and structuring tactics, than did their male counterparts in virtually every country. In Canada, the 1994 and 1998 national literacy assessments conducted by the Council of Ministers of Education, Canada (CMEC)—which embedded oracy through discussion in a writing process design—have shown that females consistently outshine males in reading and writing outcomes (CMEC, 1994, 1999). Moreover, adolescent females reported greater adoption of those scholastic and extra-scholastic processes deemed central to language arts learning by adherents of constructivist approaches. Gambell and Hunter (1999, 2000) concluded from these assessments that, by the time they reach high school, males lag behind their female counterparts in their attitudinal development, in self-concept and confidence, in frequency of reading exchange with others, in reading orally, and in discussing their writing assignments with others.

Several explanations have been offered for these persistent literacy findings, including gender proclivities for particular literary genres, divergent patterns of response to characterization in literature, and differential patterns in home socialization, among others (Gambell & Hunter, 2000). But perhaps the most pedagogically provocative model has proposed that gender differences are socially constructed as students and teachers interact in day-to-day classroom endeavour. As a generative set of hypotheses, this model's central propositions are: that gendered identities are constructed as students engage in language arts activities; that females are favoured in language acquisition through positive rewards by teachers and peers for cooperative types of learning behaviour; that teachers attend to males and females with different intents and intensities; that the gender gaps in competency, self-concept, attitude, self-efficacy, and behaviour emerge early in a student's schooling, crystallize in mid-to-late adolescence, and persist as students leave school; and that the differences in students' literacy behaviours result in higher female classroom grades, gendered results in large-scale assessments, and higher promotion and graduation rates from North American public schools (Sadker & Sadker, 1994). The central problem is whether constructivist approaches in instruction and assessment obviate girl-boy differences in their language arts development and its measured outcomes, or whether small group differences in oral production replicate those gendered outcomes found in recent literacy assessments.

Gender, group interaction and oracy

Some oral production, group interaction and gender issues have been addressed in research within constructivist pedagogy over the past two decades, with concerns around three sub-issues: whether heterogeneous groups accentuate or moderate gender differences in language acquisition; whether the type of oracy task favours one gender over another; and whether there are interrelationships between gender and learning style or self-concept as a dimension of self-identity and hence self-efficacy.

Two phases are discernable in this research. The first, roughly spanning the 1980s, coincided with the introduction of cooperative learning in classrooms as a

form of social constructive activity. As curricula upholding collaborative learning precepts were developed and distributed for use across all subject areas, researchers looked at students' performances within a variety of heterogeneous groupings in science, mathematics and computer science (DiPardo & Freedman, 1988). This curriculum-wide research interest makes it difficult to distinguish particular listening and speaking accomplishments from achievement in the disciplinary content of specific subjects. Moreover, few researchers examined within-group variables, such as cohesiveness, affective processes, or the non-verbal aspects of interaction in language arts learning groups. In the second phase during the 1990s, investigators focused on the construction of gendered identities within the language arts classroom, but primarily in relation to literature or literary tasks. That renders it difficult to partition out performance levels in particular listening and speaking skills as distinct from those activities at the service of cultivating literacy achievement. For example, Lensmire's (1994) work on small group processes in a writing workshop as an alternative learning environment, Smagorinsky and Fly's (1994) examination of teacher-led and small-group discussions with a literary anthology, and Nystrand *et al.*'s (1993) study all assume that group processes support literacy or literary outcomes.

Group composition and gender: enhancement or impediment?

Simply placing males and females together under a collaborative learning structure will not obviously ensure they will positively interact and learn (Raign & Sims, 1993). Negative interactions may reduce interpersonal engagement and actually impede the achievement. For example, peer pressures toward conformity and concurrence-seeking may result in unreflective decision-making, leading to compromises that combine the worst rather than the best of members' ideas (Battistich *et al.*, 1993). Similarly, students may adopt status-based norms that reinforce pre-existing stereotypes, so that the ideas of 'low-status' students are ignored or denigrated, while those of 'high-status' students are promoted, regardless of merit. These differences in attention and evaluation can promote 'social loafing' and foster widespread resentment among group members, both among those whose skills and ideas are not valued, and among those who believe they are doing more than their 'fair share' (Webb, 1989).

This issue of gender composition has been extensively explored in small groups of different sizes in various subjects. In science classes, for example, Madhok (1992) discovered that in majority-female groups, girls deferred to the boy; in majority-male groups, the boys ignored and insulted the girl; in all-female groups, girls had an almost equal interaction, but showed lack of confidence in understanding the experiment and difficulty even with basic procedures; in equal-male-and-female groups, status-seeking comments were low and on-task comments were high for both males and females; in all-male groups there was a wide range of turn-taking between members. Likewise, Brush (1997) found significant gender differences in total interaction and task-related help in computer-based problem solving. In

same-gender groups the female students had a higher level of total interaction, and both offered and requested more task-related help than their male counterparts. In mixed, majority-female groups, the same tendencies were detected, although they were slightly less pronounced. In equal-ratio groups the males had a higher level of interaction than did the females, and they gave more task-related help. In these groups, the females received more task-related help than the males. In majority-male groups, there was no gender difference with respect to the total interaction, but the males both asked for and gave most of the task-related help. McCaslin *et al.* (1994) found similar group processes in the mathematics classroom; all differences could be ascribed to group composition. In a language arts classroom, Maskit and Hertz-Lazarowitz (1986) found that in equal gender composition groups there was more giving of and asking for literary and language information than in groups with either gender majority. Conversely, there was a greater degree of cooperative learning in groups with either gender majority than in equal gender composition groups.

Oracy tasks and gender: group processes versus formal speaking?

Rather than focusing on the small group dynamic, some scholars have questioned the relevance of some oracy tasks and the social values placed on small group talk within the classroom. For example, Baxter (1999) asserts that female difficulties in speaking in public contexts derive from the discrepancy between what constitutes success or effectiveness in school and outside school. When collaborating in small groups, girls feel empowered by the opportunities for using an interpersonal private voice (Coates, 1995; Corson, 1997). This, it is sometimes claimed, is one of the ways in which schools privilege the discursive practices of males. Baxter argues that British schools continue to prepare boys better for speaking in the public sphere than they do girls, and what girls are good at—supportive interaction in small groups—though increasingly valorized in school settings, is not necessarily what will help them outside school, as exemplified in traditional oral English examinations (Jenkins & Cheshire, 1990, 1991).

Across the Atlantic, research typically shows North American boys dominating the public arena of the classroom, especially in teacher-directed class interactions (James & Dratich, 1993; Swann, 1989), notwithstanding the research on other gendered identities and cultural differentiation within groups that undermines such sweeping claims. Nevertheless, British researchers have emphasized two trends—the feminization of classroom talk, and the underachievement of boys. The National Oracy Project (NCC/NOP, 1991) has been prominent in British schools, promoting a model of small group, collaborative talk that has been particularly influential. ‘This has effectively produced a dominant model of classroom talk that values cooperation and consensus through informal, collaborative talk in small groups where pupils learn by discussion, by exploring an issue, by sharing ideas and viewpoints, by tackling a problem together, and by active listening’ (Baxter, 1999, p. 87). Baxter argues that this model appears to favour the cooperative discourse style of girls: thus girls are generally considered to be better at small group talk than are boys.

However, Baxter finds no statistical evidence that girls have received poor examination marks for speaking and listening compared with boys in British national English language arts examinations; rather, this is one area where boys and girls perform equally—unlike literacy, where, as in Canada and the US, girls outperform boys in reading and writing.

As for the second trend, that of boys' underachievement, Baxter argues that it is not boys who are doing worse, but girls who are doing better. She quotes from an Equal Opportunities Commission research project (Weiner *et al.*, 1996) which posits no decline in boys' oral performances; boys are doing better than ever before in examinations, and girls have simply improved even more than boys. This argument suggests that the equal opportunity reforms of the past 30 years have worked differentially, but have not led to parity in opportunity or achievement. Reflecting on the educational practices that emerged from the oracy movement in England, Swann and Graddol (1995) do not find it surprising that interactional norms became enshrined in curricula and instruction that seemed to encourage and value features associated with feminine rather than masculine speaking styles.

Interaction patterns and gender: learning styles or self-efficacy?

Is there a feminine learning style or are student learning patterns individual? Many educators claimed, prior to the debates about gender and essentialism, that learning styles vary between genders: the interactions of males are activity centred, whereas those of females are more oriented to affect and the exploration of intimate thoughts and feelings (Tannen, 1990; Belenky *et al.*, 1986; Fishman, 1978, 1980). Both males and females perceive interactions with females to be more meaningful (i.e. intimate, disclosing, pleasant, and satisfying) than interactions with males, which tend to be competitive, self-promoting, and characterized by 'one-upsmanship' (Cohen & Lotan, 1995; Reis, 1986). Goodwin (1988) has shown that females are perfectly able to switch from the more assertive and confrontational mode, such as arguing with boys or in family arguments, but choose not to be assertive with female peers when their goal is to preserve an equal and non-conflictual relationship. Girls may choose to take this same discursive path when they want to preserve a particular type of relationship with an unrelated male. O'Donnell and Smagorinsky (1999) address the question of why 'girls who appear direct, confident, and articulate in individual conferences, small groups, or in their writing, (yet) become deferential, diffident, or tentative when facing a whole class of students' (p. 35). They concluded that powerlessness, as a gendered trait, was situational: schools disadvantage females primarily when schools view knowledge as fixed and discussions are conducted to reward those who can argue their positions with the greatest certainty.

Other authors have sidestepped the essentialism and sexism controversies by focusing not on audience and situation for listening and speaking, but rather on self-identity (as revealed in a questionnaire), self-concept, and self-efficacy. Students' selection and use of learning strategies depend directly on their perceptions of their self-efficacy, whether individually or collectively, and reciprocally on the feedback

they receive. Therefore, the four efficacy-related processes (cognition, motivation, affect, and selection) work interactively to both shape students' self-perceptions of themselves as learners, their self-identity in gendered terms, and their self-confidence in school (Gist & Mitchell, 1992). Motivationally, self-efficacious learners should be able to perceive themselves as autonomous and competent, and behaviourally, they should be able to make their environment conducive to learning. Female self-efficacy is higher than boys' in reading and writing in Canada (Gambell & Hunter, 1999, 2000); a similar trend is evident internationally in reading (OECD, 2002). However, many studies have shown that female students have significantly lower self-efficacy than male students in mathematics and other traditionally male-dominated subjects, including computer science (see for example, Zimmerman & Martinez-Pons, 1990). In addition to gender, previous subject matter experience and encouragement have significantly predicted self-efficacy (Busch, 1996; Renzetti & Curran, 1999). Moreover, those gendered differences tend to solidify during a student's school years because they judge their capabilities by comparing their performances with others (Pajares, 1996).

Taken together, several conclusions can be drawn from this literature. Small group listening and speaking performances have been insufficiently explored on a large-scale basis outside the context and content of specific subject disciplines. Within small groups, the gender composition of small groups and the perceived self-efficacy of students may be important variables in learning. Small group dynamics involve not only questions of listening and speaking, but a variety of affective, kinesthetic, and other non-verbal behaviours. And investigations into the effectiveness of any small group activity must carefully articulate and distinguish between process and product, cognitive and social achievement, as well as group-expressed and individually-expressed skills and perceptions. Those presuppositions undergird the research design in the study we report here.

Method

Assumptions

To employ a social-constructive approach in the 1998 Saskatchewan Language Arts Learning Assessment (Saskatchewan Education, 1999) administered by teachers, several key premises were adopted. First and foremost, listening and speaking were conceived as an integrated, interactive act of communication, not as discrete skills. Oral communication would involve both verbal and non-verbal exchanges with an audience. The communicative content encompassed not only thoughts or information, but also feelings and attitudes. Communication was to occur in a social context: listening and speaking are means by which individuals make connections with each other. To speak fluently and confidently in a variety of situations, and to appreciate the needs and positions of others, were deemed necessary attributes of effective oral communication.

Second, the communicative task would be presented as a classroom instructional activity, not represented in pencil-and-paper and audio-electronic formats. Students

would be prompted with tasks that reflected a typical, cooperative learning response to a 'real world' situation. Rather than isolating the student in front of his or her audience or in a listening booth with earphones and cued cassette tapes, the provincial assessment asked students to watch a televised video clip and to subsequently discuss the moral and social issues raised: the qualities of people they admire, the issues in allegiances to friends, and ethical issues such as AIDS.

The third assumption was that several communicative competencies would be synthesized as primary traits in holistic rating scales in six communicative dimensions. Fully articulated criterion rubrics, with accompanying video-taped exemplar performances, were prepared involving three aspects of group discussion for assessment of listening behaviour. The first dimension was participation (such as effective approaches for organizing and completing the assignment, oral contributions that advance and stimulate discussion) which assumed that student involvement is shaped by factors such as communicative purpose, assignment focus, social structure, and language. Group members' willingness to voice ideas and opinions and to share experiences was deemed to be a part of participation, measured in terms of both the quality and quantity of participation. The second dimension was active listening skills: the verbal and non-verbal communicative behaviour of others (paraphrasing, probing and body language) were to be acknowledged and built upon. Raters had to attend to the posture and facial expression of examinees, in keeping with communicative purpose, and with the cultural and linguistic background and needs of others. And third, those attributes that demonstrated respect for conversational peers were described in rubric form. Proficient communicators convey their awareness of and concern for other members in a group setting by showing tactfulness in the expression of their ideas, polite interjections, and respect for ideas and opinions different than their own.

These presuppositions can be critically examined in terms of the way each could shape outcomes. For example, prompting students with a television or media clip may favour males, who have repeatedly reported higher levels of television watching, rather than females who generally express greater preference for more passive extra-scholastic activities such as reading. Paralinguistic and socio-communicative demonstrations of respect for the ideas and positions of others in conversation may be valued more in the classroom than in workplaces increasingly devoted to electronic communications from isolated office cubicles. Raters using traits in the active listening and participation rubrics may reward Eurocentric notions of kinesthetics and gesticulation, and misinterpret behaviours of Aboriginal and other minority students as unsupportive or inadequate. However, Aboriginal students may be advantaged in an assessment which focuses on oral traditions.

Procedure

The assessment of listening and speaking proceeded in four stages, following a blueprint somewhat different than typical, centralized large-scale standardized testing. First, after viewing a video prompt, students' individual listening abilities

were assessed using a 15-question, multiple-choice test in written form with questions read aloud to students with supporting text. This multiple-choice exercise therefore served as a measure of concurrent validity and as a pretest to an assessment of group performance that integrated active listening skills with speech communication. In the second stage, students were given a group discussion task in both oral and written form to complete within their pre-assigned small groups consisting of all-female, all-male, and mixed-gender students. They were asked to participate in a small-group discussion to prepare a 2-to-4 minute group presentation describing their collaborative response to the assignment. The student test booklet and oral instructions guided this discussion: these included preparatory questions, a planning chart and a checklist. During the small-group discussion, the pre-trained teacher rated each group for participation, active listening and respect for conversational peers. In the third phase, not reported here, each student group delivered its 2-to-4 minute presentation to their classmates with their solution to the given assignment. In the fourth stage, students completed: a self-evaluation form to rate their individual performance within the group performance; a peer evaluation form to assess how their group had performed in both the discussion and presentation phases; and a Student Questionnaire about their speaking and listening behaviours and attitudes.

Traditional pencil-and-paper tests for the three grade levels were different for each grade level, although nine questions were identical at Grade 5 and 8. The videotaped television prompts and discussion/presentation topics were different for each grade, but student questionnaires, peer evaluation forms, and teacher questionnaires were identical for participants at each grade level. When rating each group performance, Grade 5, 8 and 11 teachers used identical performance rubrics, with the same traits and attributes in each performance level phrased exactly the same for each grade level, but the videotaped exemplar performances were specific to the grade appraised. The rubrics were graduated, from Level 1 (low) through Level 4 (high), and considered to reflect stages on a continuum. The development of listening and speaking skills in a group setting was assumed to begin well before Level 1 and to extend beyond Level 4. These performance scales or rubrics holistically captured the primary characteristics of speech communication: not every attribute needed to be present in the rubric to identify a student at a particular criterion level. Identical criteria were used to rate group performance at each grade. Different examples of group performance (exemplars) were used for each grade to illustrate the type of work expected to fulfil the criteria for that grade. (For copies of rubrics and instruments, see <http://www.sasked.gov.sk.ca/k/pecs/ae/docs/plap/listenspeak/1998.pdf>)

Prior to the May 1998 assessment, teachers attended one of four, 2-day orientation sessions to develop consistency in their rating. The orientation familiarized teachers with the assessment premises and procedures, and provided intensive training for scoring student performances. During the orientation, teachers considered common rater errors, became familiar with the scoring rubrics, and practised rating with videotaped exemplar student performances using the six

scoring rubrics. Teachers were asked to share performance criteria with students before test administration, and to circulate among the groups at least twice during the assessment, observing each group performance for 5 minutes.

Reliability training to facilitate fair and reliable scoring included careful review and practice among teachers to develop a shared interpretive outlook in applying rubric criteria. Videotaped student performances were presented with a wide range of quality levels—including those that easily elicited consensus in teacher ratings, and those that provoked varied responses. During practice rounds, a scoring leader provided adjudication when necessary to enable scorers to anchor consistently in the four levels of the rubrics. During the orientation, raters checked for consistency using several indices: classic inter-rater ratios ranged from 61% to 74% on one occasion, and 61% to 75% on another. At the actual school sites, where no more than 10 teachers were double-scored, inter-rater consistency between two raters on a four-point scale ranged from 64% to 82% (Saskatchewan Education, 1999). Given the complexity of the tasks and pioneering assessment design, these conventional measures suggest moderate reliability, common in alternative assessments (Hunter & Randhawa, 2001; Nichols & Smith, 1998; Worthen, 1993).

The sample

Participating schools were randomly selected in Spring 1998 from the list of provincial schools teaching the Saskatchewan curriculum with Grade 5, 8 and 11 students. Selection criteria produced a random sample that proportionally represented rural, urban and northern (geographically dispersed, small and isolated) schools. Schools of varying populations were represented, but those controlled by federal Aboriginal Indian bands, those that had field-tested the assessment materials, those involved in parallel national or provincial reading and writing testing, and those in francophone school divisions did not participate. Schools with more than 30 students in a grade selected a class of at least 25 students.

Only those students enrolled in regular English language arts programmes were involved. Students in modified programmes, where foundational objectives for the programme had been altered, did not participate. Selected schools and their school divisions were contacted in February 1998. Information bulletins outlined the nature and scope of the assessment for teachers, parents and students. Class lists were submitted to the Department of Education for random assignment of students into small groups of four or five students for the actual assessment. This group size, rather than triads, was chosen after consideration of Saskatchewan curriculum prescriptions and typical instructional practices in Saskatchewan schools. Administrative guidelines also aimed to ensure consistent administration in terms of time allocations, breaks, independent student work, absenteeism, the type of assistance provided, and rigorous scoring of student performance. So too did the guidelines guarantee student, teacher, school, and school division anonymity. The actual assessment took place between 11 May and 22 May 1998. A total time allotment of 3½ to 4 hours was provided for the assessment in each classroom.

There were 28 schools registered in Grade 5 and 29 at each of the other two grade levels. Only one classroom from each school was identified to participate in the study. Schools submitted a class list of students to the Department of Education, from which Department officials anonymously assigned students into small groups, according to gender and the alphabetical order of the students' last names. Thus, sub-group composition crossed friendship and gender lines. A total of 608 Grade 5 students, 686 Grade 8 students, and 673 students at Grade 11 were initially identified on these classroom lists. The centrally assigned group rosters were sent back to teachers, who in a few cases (less than 2%) changed group compositions to maintain group size because of absentees. The 85 participating teachers rated a total of 551 groups of varying gender compositions over the three grade levels assessed. The actual numbers of participating students, when absentees were considered, were 577 at Grade 5, 594 at Grade 8 and 556 at Grade 11, yielding an overall response rate of 87.9%.

The assessment thus included several sources of data: a traditional multiple choice test of autonomous listening skill; teacher-assigned ratings of small group discussion involving listening and speaking in a social context; teacher-assigned ratings of a small group presentation; a student questionnaire about perceived self-efficacy in speech communication; a student self-evaluation of autonomous performance within a group; and a student evaluation of group performance in both the discussion and presentation activities. Because of the difficulties in partitioning out the gender of self-selected student presenters who represented their groups in the presentation or speaking task, these findings were excluded from analysis.

Analysis and findings

Gender differences in individual listening skills

To determine if there are gendered patterns in students' autonomous listening skills, to concurrently validate the accuracy of teacher and student ratings, and to identify potential gender disparities within a non-constructivist listening task for comparison with a constructivist listening exercise, total scores on the traditional multiple-choice listening test were analysed using an independent *t*-test. The SPSS (9.0) software for Windows was used for all the analyses reported in this paper.

The three component scores on the multiple-choice test were correlated and analysed using multivariate analysis of variance with gender as the independent variable. Multivariate *F*-tests for Grades 8 and 11 were significant with [$F(3, 590) = 2.79, p < 0.05, \text{power} = 0.67$] and [$F(3, 552) = 2.81, p < 0.05, \text{power} = 0.68$], respectively. As Table 1 shows, females tended to have raw scores higher than males in all grades and types of questions, but the differences were often small and insignificant within many dimensions. In fact, the large standard deviations show substantial variation such that many males did perform very well. Yet there were some univariate significant differences between males and females for Grade 8 Interpretation [$F(1, 592) = 5.63, p < 0.05, \text{power} = 0.82$] and Overall Performance on the entire test [$t(592) = 2.70, p < 0.05$]. At Grade 11, parallel

Table 1. Means and standard deviations for total and component scores on traditional multiple-choice listening test

Grade		Mean score			
		Comprehension	Interpretation	Analysis/evaluation	Total
5	Male ($n = 289$)	46.64 (21.73)	60.97 (18.76)	52.04 (20.72)	53.20 (14.20)
5	Female ($n = 288$)	49.10 (23.59)	61.94 (20.35)	54.72 (21.77)	59.47 (16.13)
8	Male ($n = 295$)	62.03 (19.83)	58.58 (19.22)	44.95 (21.66)	55.20 (12.53)
8	Female ($n = 299$)	65.15 (20.76)	62.27* (18.76)	46.62 (21.05)	58.00* (12.93)
11	Male ($n = 274$)	59.12 (21.70)	51.53 (21.00)	45.11 (19.89)	51.93 (13.80)
11	Female ($n = 282$)	64.26* (20.08)	52.77 (19.72)	45.89 (20.68)	54.00* (13.07)

* $p < .05$.

Note: Values in parentheses in the score columns are standard deviations.

differences emerged for Comprehension [$F(1, 554) = 8.38, p < 0.05, \text{power} = 0.82$], and Overall Performance [$t(554) = 2.09, p < 0.05$]. None of the gender differences were significant for Grade 5 students.

Expressed in terms of student performance, then, male and female students in Grade 5 classrooms showed no significant differences in their independent ability to comprehend, interpret or evaluate linguistic and psycholinguistic cues derived from a video clip dealing with ethical issues that involved theft of materials. By Grade 8, however, with the same videotaped prompt, some differences favouring females were apparent in autonomous ability to make interpretations of those cues, and overall in the total independent listening act, although scores for the comprehension and analysis subcomponents were not significantly different. This modest female superiority persisted in Grade 11, as demonstrated in relation to a parallel videotaped listening task, and particularly in the independent ability to comprehend audio cues. Yet there were no significant differences between high school males and females in their interpretive and analytic or evaluative skills when listening to the same material. Grades 8 and 11 females performed best on comprehension questions and lowest on analysis-evaluation questions, as did males. Both Grade 5 females and males had the most difficulty with comprehension questions. The Grades 5 and 8 multiple-choice tests included nine questions in common. A larger percentage of Grade 8 students than Grade 5 students answered each of the nine common questions correctly, suggesting that students' depth of understanding does increase as they grew older, regardless of gender. In sum, female Saskatchewan students appeared to become modestly superior in their overall, autonomous listening skills of comprehension and interpretation as they moved into secondary school.

Group composition and quality of group discussion

To explore gender interaction patterns, teacher ratings for the three dimensions of group performance in discussion were used as the dependent variables, and a

three-factor multivariate analysis of variance was conducted. The first factor was Grade (L), a crossed factor with three levels: Level 1 = Grade 5, Level 2 = Grade 8 and Level 3 = Grade 11. The second factor was School (S), a nested factor within L, where there were 28 at the Grade 5 level and 29 schools for each of the other two grades. The third factor was Group Composition (G), a nested factor within S that was categorized into five levels: all male (5), majority male (4), equal gender (3), majority female (2), and all female groups (1). Multivariate significant effects were found for grade level L [$F(6, 1064) = 2.90, p < 0.05, \text{power} = 0.90$] and G [$F(12, 1595) = 2.07, p < 0.05, \text{power} = 0.94$]. The only corresponding univariate significant effect for L was found for the Respect for Conversational Peers dependent variable. However, for the G factor, all three univariate effects were significant, indicating that there were differences in the rated performances of each of the three dimensions for the five types of group memberships. In order to discern the sources of univariate significant differences, the Scheffe pair-wise comparison option in the SPSS program for the L and G factors was employed.

Table 2 presents the combined means and standard deviations of the three dimensions for the L and G factors. Table 2 also shows the homogeneous sets within a dependent variable for the L and G factors. Means that do not share the same superscript are significantly different from each other. For example, when examining grade level (L) means for the Respect for Conversational Peers dimension, Grade 5 has superscript 1, Grade 8 likewise has superscript 1, and Grade 11 has only 2 for its superscript. To summarize in terms of grade level performance, on the Active Listening and Respect for Conversational Peers, Grade 11 participants scored significantly higher ($p < 0.05$) than the Grade 5 and Grade 8 participants,

Table 2. Combined means and standard deviations on three group discussion dimensions by grade and group composition

Independent variable		<i>n</i>	Group discussion dimension		
			Participation	Listening	Respect
Grade (L)	5	177	2.51 (0.83)	2.61 ¹ (0.76)	2.66 ¹ (0.85)
	8	193	2.58 (0.83)	2.59 ¹ (0.87)	2.77 ¹ (0.85)
	11	180	2.69 (0.79)	2.83 ² (0.82)	2.95 ² (0.69)
Group Composition (G)	All female	69	2.87 ¹ (0.82)	3.01 ¹ (0.92)	3.03 (0.89)
	Majority female	180	2.64 ^{1,2} (0.82)	2.73 ^{1,2} (0.81)	2.89 (0.84)
	Equal male–female	37	2.54 ^{1,2} (0.84)	2.62 ^{1,2} (0.72)	2.65 (0.79)
	Majority male	210	2.53 ^{1,2} (0.78)	2.59 ² (0.78)	2.74 (0.76)
	All male	54	2.37 ² (0.83)	2.39 ² (0.86)	2.67 (0.78)
Total		550	2.59 (0.82)	2.67 (0.83)	2.81 (0.81)

^{1,2}Superscripted numerals beside a dependent variable indicate homogeneous sets for independent variable categories, Grade or Group Composition.

n = number of student groups rated by teacher scorers.

Note: Values in parentheses are standard deviations.

respectively. However, the performances of Grades 5 and 8 participants on these two dimensions were not statistically different.

But as Table 2 reveals, some differences emerged when one considers the gender composition of groups combined across three grades in the Saskatchewan assessment. Groups varied from homogenous all-female to all-male in membership: the all-female and majority-female groups exceeded average group performance on all three dimensions of group performance when rated by teachers (Participation, Active Listening, Respect for Peers). And inversely, for all three dimensions, majority-male and all-male groups scored below average in their group performance. Below average results were also the norm for equal-female/male, majority-male and all-male groups at each grade for each rubric deployed. Although, as noted earlier, significant univariate effects for group composition (G) emerged for all three dimensions of group discussion, pair-wise significant differences were noted for Participation and Active Listening. On the Participation variable, the all-female groups had a significantly higher mean than the all-male groups. No other pair-wise comparison was significant. Only on the Active Listening dimension did the all-female groups have a significantly higher mean than both the majority-male and all-male groups.

Expressed another way, only homogenous group composition appears generally as a significant variable. All-female groups consistently and significantly scored more highly than all-male groups across the total assessment, but not heterogenous groups. The performance differences were generally insignificant among majority-female, equal-male/female and majority-male groups. In general, when positioned within a continuum that embraces single gender groupings, there is a consistent gradient in scores; the more heavily weighted the group in female membership, the higher the rated group performance in making perceptive insightful comments and working purposively and harmoniously, in using verbal and non-verbal communicative strategies to become actively involved in listening, and in demonstrating respect for the ideas and opinions of others. But it was only the all-male small groups that were significantly less effective in oral production and discursive practice than all-female groups.

Gender and self-efficacy

Do boys and girls hold differing beliefs about their own effectiveness in listening and speaking? Five student questionnaire items, involving self-efficacy in discursive practice shown in Table 3, were analysed using a two-fixed factor MANOVA. Cross-grade comparisons can be reliably made because all students completed a common instrument. The first factor was Gender (A) with two levels, male and female, and the second factor was Grade (L) with three levels, Grades 5, 8, and 11. All effects were significant: A [$F(5, 1732) = 16.49, p < 0.05, \text{power} = 1.0$]; L [$F(10, 3464) = 8.85, p < 0.05, \text{power} = 1.0$]; and interaction $A \times L$ [$F(10, 3464) = 2.77, p < 0.05, \text{power} = 0.97$]. The corresponding univariate results indicated that, for all five items, there were significant gender differences such that, overall, female

Table 3. Means and standard deviations^a in student self-efficacy, by gender and grade

	<i>n</i>	Grade 5		Grade 8		Grade 11	
		M	F	M	F	M	F
Self-efficacy dimension		285	297	292	305	273	290
1. Importance of good listening skills in adult life		3.77(0.66)	3.81(0.66)	3.67(0.72)	3.68(0.60)	3.58(0.70)	3.82*(0.43)
2. Importance of effective speech communication in adult life		3.78(0.90)	3.94*(0.90)	3.60(0.88)	3.72(0.72)	3.52(0.77)	3.80*(0.47)
3. Self-perception as listener		3.54(0.80)	3.75*(0.76)	3.67(0.87)	3.79(0.87)	3.74(0.83)	4.05*(0.74)
4. Self-perception as speaker		3.52(0.90)	3.66(0.88)	3.63(0.86)	3.60(0.91)	3.38(0.92)	3.70*(0.81)
5. Self-reported persistence in completing quality work		3.60(1.05)	3.76(1.10)	3.41(0.93)	3.69*(0.86)	3.21(0.98)	3.69*(0.81)

* $p \leq 0.05$.

^aValues not in parentheses are means of self-reported student ratings on a five-point Likert scale, 1 = low and 5 = high. Values in parentheses are standard deviations.

n=number of students completing questionnaire at a particular grade level by gender.

students demonstrated higher self-efficacy than males. On the Grade factor, the univariate results were significant for all items except Item 4: Self-perception as speaker. The interaction univariate results were significant for three items: (1) Importance of good listening; (2) Self-perception as speaker; and (3) Self-reported persistence. Generally, these results suggest that differences in self-efficacy depend upon both the gender and the grade level of the students.

Post hoc pair-wise multiple comparisons of the grade factor, using the Scheffe procedure, revealed that Grade 11 students had significantly ($p < 0.05$) lower levels of perceived self-efficacy than Grade 5 students on 'Importance of effective speech' and 'Self-reported persistence'. However, Grade 11 students demonstrated higher self-efficacy than Grade 5 students on 'Self-perception as listener'. Also, Grade 11 students had significantly ($p < 0.05$) higher self-efficacy perceptions than Grade 8 students on 'Self-perception as listener'. Furthermore, Grade 8 students had significantly ($p < 0.05$) lower means than Grade 5 students on two items: 'Importance of good listening' and 'Importance of effective speech'. More generally, male attitudes slowly but consistently declined from Grade 5 to 8 to 11 in their views of the importance of listening skills in adult life, of effective speech communication's importance in adult life, and in their self-reported persistence in completing quality oral work in the classroom. Conversely, male self-perceptions as effective listeners improved consistently as they grew older. Similarly, female students' self-perceptions as listeners improved as they grew older. For all other dimensions, female self-efficacy declined from Grade 5 to 8 but reverted to approximately its original level by the time they were in Grade 11.

Thus, as shown in Table 3, the gender differences were statistically significant ($p < 0.05$) for all the variables of Grade 11 self-efficacy reviewed, but for only two of the Grade 5 variables and one at Grade 8. In most aspects of self-efficacy—whether in students' estimation of effective listening skills and speech communication in adult life, in their self-defined capacity as a listener or speaker, or in their self-described persistence in completed oral tasks in the classroom to a high quality—female students reported greater efficacy than males. The exceptions where self-efficacy was roughly equal for boys and girls were in Grade 8 estimation of listening skills' importance in the adult world and in self-perception as speakers. A significant girl-boy gap in self-efficacy had become apparent at Grade 11 that was not generally evident at early grades.

Gender patterns in self- and group-evaluation

If there were gender differences in self-efficacy, or in individual's beliefs about their general effectiveness with a skill, were there parallel differences in perceived ability to execute the specific task at hand? How well did students evaluate their own and their group's skills on the particular group discussion task before them? Table 4 presents means and standard deviations of items used for student evaluation of their own performance (self) and also their peer group during the group discussion. The same questionnaire items were used for Grades 5, 8, and 11 assessment; however, the

Table 4. Means and standard deviations^a of student self-evaluation and student evaluation of group performance, by gender and grade

	Grade 5		Grade 8		Grade 11	
	M	F	M	F	M	F
<i>Student self-evaluation</i>						
1. I offered ideas and suggestions	3.08(0.72)	3.15(0.70)	3.16(0.72)	3.16(0.71)	3.21(0.65)	3.35*(0.61)
2. I helped group stay focused	2.88(0.81)	3.02*(0.81)	2.78(0.78)	2.85(0.72)	2.90(0.73)	2.94(0.72)
3. I listened carefully and courteously	3.42(0.65)	3.39(0.77)	3.39(0.63)	3.41(0.66)	3.42(0.70)	3.48(0.59)
4. I asked questions when not understood	2.85(0.96)	3.03*(0.92)	2.99(0.92)	3.07(0.80)	2.99(0.80)	3.10(0.77)
5. I encouraged members to share ideas	2.79(0.94)	2.90 (0.92)	2.75(0.90)	2.91*(0.86)	2.93(0.83)	3.10*(0.77)
6. I helped group reach consensus	3.17 (0.81)	3.20(0.80)	3.22(0.75)	3.27(0.73)	3.15(0.77)	3.35*(0.63)
<i>Student evaluation of group performance</i>						
1. Offered ideas and suggestions and shared responsibility	3.47(0.67)	3.44 (0.70)	3.44(0.63)	3.40(0.70)	3.47(0.64)	3.52(0.63)
2. Stayed on topic and worked well together	3.10(0.78)	3.18(0.81)	3.11 (0.71)	3.04(0.83)	3.20(0.74)	3.20(0.71)
3. Listened carefully and courteously	3.31(0.74)	3.30(0.76)	3.40(0.62)	3.37(0.69)	3.40(0.68)	3.54*(0.63)
4. Asked questions to gather information and seek clarity	3.02(0.84)	3.04(0.83)	3.07(0.74)	3.15(0.77)	3.21(0.72)	3.31(0.69)
5. Showed respect for each other	3.41(0.75)	3.46(0.77)	3.38(0.74)	3.40(0.75)	3.55 (0.70)	3.63(0.65)
6. Valued and understood ideas of other members	3.24(0.73)	3.28(0.69)	3.36(0.62)	3.28(0.74)	3.43(0.62)	3.48(0.66)

* $p < 0.05$.^aStandard deviations are in parentheses.

tasks for discussion and preparation for presentation by students in these grades were different. Therefore, the statistics in Table 4 were produced by gender at each grade level, and an independent *t*-test was employed to establish if gender differences were significant.

Results show that out of the 12 questionnaire items, on only two at Grade 5, one at Grade 8, and four at Grade 11, were there significant ($p < 0.05$) differences between males' and females' mean scores, each favouring females. At Grade 5 and 8 levels the significant gender difference could be attributed to chance; however, at Grade 11 the observed significant gender differences and the trend in differences suggest a crystallization of gender differences. At this grade level, in all except one, where the observed means were equal, females had higher means than their male counterparts. Such a consistent pattern of observed differences was not found at Grades 5 and 8.

Discussion and implications

The 1998 Saskatchewan learning assessment of students' speech communication skills in a social context yielded many significant gender differences in individual listening skills, in group production, and in self-efficacy in secondary schools. Adolescent male students were disadvantaged in the basic cognitive skills of understanding, interpreting and analysing linguistic content from visual formats through auditory modes. Moreover, when asked to actively engage in cooperative oral language processes, small groups of majority- or all-male groups lagged significantly behind female groups in a secondary school setting.

Viewed in terms of the group traits assessed, all-female or majority-female groups used more effective participatory approaches for organizing and completing the oral language assignment, made better oral contributions that advanced and stimulated discussion, and were more willing to voice ideas and opinions and to share experiences, both in quality and quantity. In groups composed predominately or exclusively of females, students showed better active listening tactics, including those verbal and non-verbal communicative tactics of paraphrasing, probing and using appropriate body language. Adolescent female-dominant groups also manifested greater respect for conversational peers in conveying their awareness of and concern for other group members by showing tactfulness in the expression of their ideas, polite interjections, and respect for ideas and opinions different than their own.

None of these girl-boy gaps in speech communication within small groups were apparent at the elementary grade assessed, but were evident by Grade 8 and largely persistent at Grade 11. In parallel fashion, a gender gap in the cumulative skills of comprehending, interpreting and analysing through aural modes had emerged in early adolescence that did not shrink or close by late adolescence. Adolescent males were disadvantaged by the time they reached high school in their estimation of the importance of listening and speaking skills, in their self-perception of linguistic effectiveness, in the way they made meaning with others, in oral production, and in discussing their language assignments with others, notwithstanding improved

self-perceptions as listeners. Whereas female attitudinal development in the oral modes appears arrested between ages 10 to 13, it recovers, while for males, it is arrested or erodes between ages 10 to 16. These Saskatchewan findings thus corroborate studies that show adolescent females as more readily adopting those scholastic and extra-scholastic processes deemed central to language arts learning by adherents of constructivist approaches.

The girl–boy gaps in this study cannot be directly linked to particular instructional approaches, since no correlational analyses were made to link the teaching methods of teachers in the regular classroom context with group performance on this assessment. However, the large-scale assessment itself enacted goals found in provincial (state) curriculum guides based on constructivist assumptions, and did not just assess student skills. Authentic assessment proponents claim to both model and measure student behaviour. An explicitly constructivist assessment yielded significant gender gaps in performance and perception, indicating that constructivist approaches may not be generating equitable opportunities for males and females.

The interrelationships between literacy and oracy remain opaque in this study, because of the assessment's design. The evident girl–boy gaps cannot be attributed to differential approaches to engaging with literary text, since the assessment deployed verbal and video prompts. Gender differentials emerged in the preparation of a group presentation on a social or ethical issue, not in relation to making meaning with reading and writing print materials. We can say that adolescents' oral communication replicates gendered differences in adolescent literacy, but we cannot affirm that it accentuates or attenuates the disparity in written production or reading competence. Nor can we say that the gendered differences in group performance cause a girl–boy gap in achievement as measured in a group product. In this project, what was measured was group production, not the outcome of small group activity (Webb, 1995). Although a subsequent group presentation did show similar gender gaps to those in the preceding group discussion (Saskatchewan Education, 1999), no analyses were made of relationships between the small group discussion and subsequent group presentation.

Although this study provides limited insight into specifically where the gender differences arise in the classroom confluence between student ability (Todd, 1992), teacher interventions (Cohen & Lotan, 1995), peer interactions (Perrenet & Terwel, 1997; Webb, 1989), and task, it is clear that Saskatchewan female teens demonstrate superiority in those independent listening, team work, participatory, and cooperative behaviours increasingly valued in Canadian language arts curricula and, at least rhetorically, in the Canadian workplace. Studies over the last 20 years increasingly point to the need for collaborative teamwork, the ability to think critically, and the ability to be creative in problem solving, as central for workforce effectiveness, despite the systemic advantages accorded males. Therefore, female superiority in these skills and abilities does not seem to transpose to the workplace. This disjunction between socio-linguistic skills developed and valued in schools and those valued in workplaces is the point made by Jenkins and Cheshire (1990, 1991) earlier in this article.

Girl-boy gaps in self-efficacy, not apparent in the elementary and middle years, become significant at the secondary level. In students' estimation of effective listening skills and speech communication in adult life, in their self-defined capacity as a listener or speaker, or in their self-described persistence in completing oral tasks in the classroom to a high quality, adolescent female students reported greater efficacy than did males. This study thus echoes national (CMEC, 1994, 1999) and international studies (OECD, 2002) that show identical gaps between males and females in reading self-efficaciousness. A causal relationship between self-efficacy and group discussion performance cannot be concluded from these data, but can be inferred. Yet the complex and probably cumulative interrelationships between self-efficacy and motivation, affect and self-control in social situations suggest that small group activities are reinforcing, not remediating, males' propensity to underperform in the high school English language arts classroom.

Some literacy researchers have posited that the gender imbalances arise from the task, speculating that males demonstrate superiority when the communicative purposes are expository, functional and pragmatic, whereas females show greater proficiency in narrative, aesthetic and imaginative situations (Gambell & Hunter, 1999). One might argue that with tasks requiring a solution or resolution, the manner in which the group arrives at an acceptable conclusion matters less than the product. In the Saskatchewan study, the task situation had immediate and pragmatic purposes and procedures, the purposes were external, and the audience prescribed was teacher and peers. Would group discussion among boys have been different had they greater freedom in pursuing their own purpose, topic, and/or audience? The task itself tended to elicit verbal analysis, exposition, and ethical consideration rather than narration and aesthetic response. Viewed as such, the findings might undermine the claim that adolescent males will demonstrate greater proficiency in practical and functional situations.

In general, the findings support the conclusions of cooperative learning researchers who have found that group composition is a significant variable in performance. Unlike other studies, the Saskatchewan study did not show that heterogeneous groups score either lower or higher than homogenous groups of either gender (Brush, 1997; McCaslin *et al.*, 1994), but rather that homogenous groups of females will tend to score significantly better than homogenous groups of males in listening and speaking at the secondary level. Random teacher assignment of, or peer self-selected membership in, groups may inadvertently construct gender disparities; the proclivity of females to cooperate with teachers and other learners in the classroom is another plausible explanation. All-male or majority-male groups may actually create a critical mass where boys can subvert the teacher-set purposes, topics, and procedures for learning. To address male/female gaps in literacy, some have posited that all-male classes can be a remedy (Bushweller, 1994); similar arguments have been advanced for all-female classes and schools (Richards, 1980). The Saskatchewan findings suggest pause for further thought before entertaining this simplistic solution.

The gender gap also raises questions about classroom evaluation practice, and particularly about the trustworthiness of holistic approaches to scoring. Critics have suggested that, in classroom evaluation of writing, markers may be responding to the cosmetic features of a student performance, such as neat handwriting or standardized spelling, rather than the substantive (Charney, 1984). Some have found a halo effect in teacher-awarded scores in one Canadian province across most high school subjects, advantaging females (Pope *et al.*, 2002). Others (Roen, 1992; Peterson, 1998) have found an interaction between the marker's gender and the student's gender, with scorers favouring writing by students of their own gender and privileging the narrative modes of girls. Does the same hold true when teachers appraise students' speech communication in a social setting? What distinguishes the superficial from the substantive in oral production? With the increasing inclusion of analytic scoring, primary trait and holistic criterion rubrics in language arts curricula and supporting resources, we wonder whether teachers have sufficient orientation in their unbiased use.

Conclusions

Because of its centrality to school success, social status, and workplace effectiveness (Coates, 1995), oral and aural skills development has been increasingly emphasized in Canadian curricula, primarily in language arts but also in other subject areas. In tandem, authentic assessors have touted open-response formats and complex measures for evaluating student work, as a lever for standards-based school reform. In doing so, both curriculum writers and programme evaluators have sought to elicit and measure a broader range of student performances. Their aim has been to enable all students to both learn and exhibit their learning in meaningful ways within provincial and pan-Canadian agendas to improve and equalize educational outcomes for students.

The gendered outcomes in public school literacy practices and products favouring females across Canada and the US since at least the early 1990s may be paralleled and preceded by gendered processes in speaking and listening. To date we know of no other large-scale oracy assessments with the array of data already available nationally and internationally for school literacy through multiple assessments over time. Hence, without a baseline, it remains difficult to ascertain whether social constructivist curricula are amplifying or moderating gender gaps in school oracy. But the 1998 Saskatchewan listening and speaking assessment findings do suggest that we must question underlying presumptions of gender equity as curricula, instructional approaches, and assessment practices are revised. A fundamental issue is whether constructivist curricular and assessment approaches enable better teaching, create a more equitable environment for learning, transform schooling for under-served students, or whether they merely reify existing inequities.

Constructivists often assume that group approaches will automatically create a mutually beneficial learning environment for all members, regardless of gender. However, group processes need to be modelled and taught; students need to know

how to make one's point known without denigrating the contributions of others, how to foster consensus in a climate of controversy. The socio-linguistic dynamics of group interaction may not yet be sufficiently understood by teachers and students, especially in mixed gender groupings. Indeed, the principal message from this study may be that gender differences persist, even when evaluative procedures are expanded to encompass gestures, body language, posture and other kinesthetic and paralinguistic traits.

This large-scale oracy assessment points then to several issues worthy of further exploration. Of particular interest would be micro-ethnographic exploration of group language interactions in those critical early adolescent years between Grades 5 and 8, because this large-scale project detected a dramatic gender shift at Grade 8. The almost exclusive focus on school and home literacy by researchers in the last decade needs to be complemented by a sustained research and applied development effort in oracy. Although some have conjectured that gesticulation is at the leading edge of children's cognitive development (see for example, Roth, 2001), little research has focused on its gendered dimensions in oral communication, except in the most general ways. Educational drama (as opposed to theatre) has been incorporated in high school language arts curricula, but may not be operationalized in ways that explore the many ways in which the body and voice interact to create meaning and interpretation, and to enact knowing.

Moreover, we know virtually nothing about how classroom teachers generate report card marks for the oral modes and about the weight they assign listening and speaking activities in their overall evaluations. Whereas aural and oral production has a central place in teaching second language courses, it is often relegated to lesser status in first language classes at the high school level. Focused attention needs to be paid to developing rubrics and exemplars as devices for articulating expectations and standards not only for literacy but for listening and speaking and for group processes and presentations.

Our study also makes a case for reconsidering the position of listening and speaking development within curriculum and not just evaluative practice. Too often, collaborative group activity is subordinated to literacy activities and literary content, when such cooperative learning strategies as jigsaw activities deserve deliberate attention in their own right. Read-aloud protocols, small group activities, and formal class presentations now serve primarily as levers for improving reading and written production. The pertinent skills of listening and speaking are too often taught only insofar as they improve a student's understanding of print text, as vehicles for appreciating and improving reading, and as prompts for enhancing written expressiveness. Social constructive ideals have led to reading and writing across curricula, without any reciprocal effort to mesh social studies goals with the language arts curriculum. We wonder if some learning outcomes generally assigned to social studies teachers—how to read history as narrative, communicative etiquette in mixed status groups, effective expression of dissent, the use of persuasion, structuring a verbal argument in ethically responsible ways, recognizing and responsibly critiquing stereotypes, attending to and demonstrating respect for

culturally different non-verbal cues in body language and gesture, exercising democratic notions of power, critique and consensus in groups—deserve prominence in language arts frameworks.

Finally, our study suggests that teachers should address gender differences, not through segregating but by directly attending to the dynamics by which both gender and meaning are constructed in the small group. When we educators adopt a *laissez faire* approach to small group processes, pass over kinesthetic non-engagement, or ignore a cutting gendered remark, we are not only condoning small group dysfunction and undermining gender equity; we are also impeding the development of appropriate social skills for adult life, undermining the cultivation of self-efficacy, and detracting from the creation of a literate culture itself. Because oral language occupies a central position in the classroom, its use must be planned and pondered if we are to constructively address gender issues in language arts and education in general.

Note

1. In Canada, 'language arts' curriculum refers to the teaching of English at all grade levels. Another common descriptor is 'English language arts' as different from, say French language arts. Throughout this paper we use the term 'language arts' to refer to all grade levels where English is the subject matter.

References

- Aries, E. (1976) Interactions, patterns and themes of male, female and mixed groups, *Small Group Behavior*, 7, 7–13.
- Aries, E. (1977) Male–female interpersonal styles in all male, all female, and mixed groups, in: Alice Sargent (Ed.) *Beyond sex roles* (St Paul, MN, West Publishing).
- Bandura, A. (1993) Perceived self-efficacy in cognitive development and functioning, *Educational Psychologist*, 28(2), 117–148.
- Bandura, A. (1995) Exercise of personal and collective efficacy in changing societies, in: A. Bandura (Ed.) *Self-efficacy in changing societies* (New York, Cambridge University Press), 1–45.
- Bandura, A. (1997) *Self-efficacy: the exercise of control* (New York, Freeman).
- Barnes, D. & Todd, F. (1977) *Communication and learning in small groups* (London, Routledge).
- Baxter, J. (1999) Teaching girls to speak out: the female voice in public contexts, *Language and Education*, 13(2), 81–98.
- Battistich, V., Solomon, D. & Delucchi, K. (1993) Interaction processes and student outcomes in cooperative learning groups, *The Elementary School Journal*, 4(1), 19–32.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R. & Tarule, J. M. (1986) *Women's ways of knowing* (New York, Basic Books).
- Brush, T. (1997) The effects of group composition on achievement and time on task for students completing ILS activities in cooperative pairs, *Journal of Research on Computing in Education*, 30(1), 1–17.
- Busch, T. (1996) Gender, group composition, cooperation, and self-efficacy in computer studies, *Computer Studies*, 15(2), 25–135.
- Bushweller, K. (1994) Turning our backs on the boys, *American School Board Journal*, 181(5), 29–35.
- Charney, D. (1984) The validity of using holistic scoring to evaluate writing: a critical review, *Research in the Teaching of English*, 18(1), 65–81.

- Coates, J. (1995) Language, gender and career, in: S. Mills (Ed.) *Language and gender* (London, Longman).
- Cohen, E. & Lotan, R. (1995) Producing equal-status interaction in the heterogeneous classroom, *American Educational Research Journal*, 2(1), 99–120.
- Cook-Sather, A. (2002) Authorizing students' perspectives: toward trust, dialogue, and change in education, *Educational Researcher*, 31(4), 3–14.
- Corson, D. (1997) Gender, discourse and senior education: ligatures for girls, options for boys?, in: R. Wodak (Ed.) *Gender and discourse* (London, Sage).
- Council of Ministers of Education, Canada (CMEC) (1994) *School achievement indicators program: 1994 reading and writing assessment* (Toronto, CMEC).
- Council of Ministers of Education, Canada (CMEC) (1999) *School achievement indicators program: 1998 reading and writing II assessment* (Toronto, CMEC).
- Department of Education and Science (DES) (1975) *A language for life* Report of the Committee of Inquiry of the Secretary of State for Education and Science under the Chairmanship of Sir Alan Bullock (London, HMSO).
- DiPardo, P. & Freedman, S. (1988) Peer response groups in the writing classroom: theoretical foundations and new directions, *Review of Educational Research*, 58(2), 119–149.
- Fishman, P. (1978) Interaction: the work women do, *Social Problems*, 42, 397–406.
- Fishman, P. (1980) Conversational insecurity, in: H. Giles, W. P. Robinson & P. Smith (Eds) *Language: social psychological perspectives* (Oxford, Pergamon), 127–132.
- Gambell, T. & Hunter, D. (1999) Rethinking gender differences in literacy, *Canadian Journal of Education*, 24(1), 1–16.
- Gambell, T. & Hunter, D. (2000) Surveying gender differences in Canadian school literacy, *Journal of Curriculum Studies*, 32(5), 689–719.
- Gist, M. E. & Mitchell, T. R. (1992) Self-efficacy: a theoretical analysis of its determinants and malleability, *Academy of Management Review*, 17, 211.
- Goodwin, M. (1988) Co-operation and competition across girls' play activities, in: J. Coates (Ed.) *Language and gender: a reader* (Oxford, Blackwell), 121–146.
- Governments of Alberta, British Columbia, Manitoba, North-west Territories, Saskatchewan, and Yukon Territory (1998) *The common curriculum framework for English language arts: Kindergarten to Grade 12*. Western Canadian Protocol for Collaboration in Basic Education (Edmonton, AB, Victoria, BC, Winnipeg, MB, Yellowknife, NT, Regina, SK, Whitehorse, YT, Governments of Alberta, British Columbia, Manitoba, North-west Territories, Saskatchewan, and Yukon Territory).
- Hunter, D. M. & Randhawa, B. S. (2001) The large-scale, authentic assessment of listening and speaking: issues in reliability, *Alberta Journal of Educational Research*, XLVII(2), 156–173.
- Hynds, S. (1997) *On the brink: negotiating literature and life with adolescents* (New York, Teachers College Press).
- James, D. & Dratich, J. (1993) Understanding gender differences in amount of talk: a critical review of the research, in: D. Tannen (Ed.) *Gender and conversational interaction* (New York, Oxford University Press), 281–312.
- Jenkins, N. & Cheshire, J. (1990) Gender issues in the GCSE Oral English Examination: Part I, *Language and Education*, 4(4), 261–292.
- Jenkins, N. & Cheshire, J. (1991) Gender issues in the GCSE Oral English Examination: Part II, *Language and Education*, 5(1), 19–40.
- Lee, C. & Bobko, P. (1994) Self-efficacy beliefs: comparison of five measures, *Journal of Applied Psychology*, 79, 364–369.
- Lensmire, T. (1994) Writing workshop as a carnival: reflections on an alternative learning environment, *Harvard Educational Review*, 64(4), 371–391.
- Madhok, J. J. (1992) The effect of gender composition on group interaction, in: K. Hall, M. Bucholtz & B. Moonwoman (Eds) *Locating power*, proceedings of the 2nd Berkeley

- Women and Language Conference* (vol. 2), Berkeley, CA, 4–5 April 1991 (Berkeley, CA: Berkeley Women and Language Group, University of Berkeley), 371–386.
- Maskit, D. & Hertz-Lazarowitz, R. (1986) Adults in cooperative learning: effects of group size and group gender composition on group learning, paper presented at the *Annual Meeting of the American Educational Research Association*, San Francisco, CA, ERIC: ED279788.
- McCaslin, M., Tuck, D., Wiard, A., Brown, B., LaPage, J. & Pyle, J. (1994) Gender composition and small-group learning in fourth-grade mathematics, *The Elementary School Journal*, 94, 467–482.
- National Curriculum Council/National Oracy Project (NCC/NOP) (1991) *Teaching, talking and learning at Key Stage 3* (York, National Curriculum Council).
- Nichols, P. D. & Smith, P. L. (1998) Contextualizing the interpretation of reliability data, *Educational Measurement, Issues and Practice*, 17(3), 24–36.
- Nystrand, M., Gamoran, A. & Heck, M. J. (1993) Using small groups for response to and thinking about literature, *English Journal*, 82(1), 14–22.
- O'Donnell, C. & Smagorinsky, P. (1999) Revising Ophelia: rethinking questions of gender and power in school, *English Journal*, 88(3), 35–42.
- Organization of Economic Cooperation and Development (OECD) (2002) *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000* (Paris: OECD).
- Pajares, F. (1996) Self-efficacy in academic settings, *Review of Educational Research*, 66(4), 543–578.
- Perrenet, J. & Terwel, J. (1997, March) Interaction patterns in cooperative groups: The effects of gender, ethnicity and ability. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Peterson, S. (1998) Evaluation and teachers' perceptions of gender in sixth-grade student writing, *Research in the Teaching of English*, 33(2), 181–208.
- Pope, G., Wentzel, C. & Cammaert, R. (2002) Relationships between gender and Alberta diploma scores, *Alberta Journal of Educational Research*, XLVIII(4), 275–286.
- Raign, K. & Sims, B. (1993) Gender, persuasion techniques, and collaboration, *Technical Communication Quarterly*, 2(1), 89–104.
- Reis, H. T. (1986) Gender effects in social participation: intimacy, loneliness, and the conduct of social interaction, in: R. Gilmour & S. Duck (Eds) *The merging field of personal relationships* (Hillsdale, NJ, Lawrence Erlbaum), 91–105.
- Renzetti, C. & Curran, D. (1999) *Women, men, and society* (4th edn) (Boston, MA, Allyn & Bacon).
- Richards, J. (1980) *The sceptical feminist: a philosophic inquiry* (Boston, MA, Routledge & Kegan Paul).
- Roen, D. (1992) Gender and teacher response to student writing, in: N. McCracken & B. Appleby (Eds) *Gender issues in the teaching of English* (Portsmouth, NH, Heinemann), 126–141.
- Roth, W. M. (2001) Gestures: their role in teaching and learning, *Review of Educational Research*, 7(3), 365–392.
- Sadker, M. & Sadker, D. (1994) *Failing at fairness: how America's schools cheat girls* (New York, Charles Scribner's).
- Saskatchewan Education (1999) *The 1998 provincial learning assessment in English language arts (listening and speaking)* (Regina, SK, Saskatchewan Education) (For instruments and rubrics, see <http://www.sasked.gov.sk.ca/k/pecs/ac/docs/plap/listenspeak/1998.pdf>).
- Smagorinsky, P. & Fly, P. K. (1994) A new perspective on why small groups do and don't work, *English Journal*, 83(3), 54–58.
- Swann, J. (1989) Talk control: an illustration from the classroom of problems in analysing male dominance of conversation, in: J. Coates & D. Cameron (Eds) *Women in their speech communities* (Harlow, NY, Longman), 123–140.
- Swann, J. & Graddol, D. (1995) Feminising classroom talk?, in: S. Mills (Ed.) *Language and gender: interdisciplinary perspectives* (London and New York, Longman), 135–148.

- Tannen, D. (1990) *You just don't understand: women and men in conversation* (New York, Ballantine Books).
- Todd, T. (1992) Interactive listening: an examination of listening ability and gender differences in an interactive conversational context. Paper presented at the 13th Annual Meeting of the International Listening Association, Seattle, WA.
- Webb, N. M. (1989) Peer interaction and learning in small groups, *International Journal of Educational Research*, 13, 21–39.
- Webb, N. (1995) Group collaboration in assessment: multiple objectives, processes, and outcomes, *Educational Evaluation and Policy Analysis*, 17(2), 239–1.
- Weiner, G., Arnot, M. & David, M. (1996) *Education reforms and gender equality in schools* (London, Equal Opportunities Commission).
- Worthen, B. R. (1993) Critical issues that will determine the future of alternative assessment, *Phi Delta Kappan*, 74, 444–454.
- Zimmerman, B. (1995) Self-efficacy and educational development, in: A. Bandura (Ed.) *Self-efficacy in changing societies* (New York, Cambridge University Press), 202–231.
- Zimmerman, B. & Martinez-Pons, M. (1990) Student differences in self-regulated learning: relating grade, sex, and giftedness to self-efficacy and strategy use, *Journal of Educational Psychology*, 82(1), 51–59.

Copyright of Educational Review is the property of Educational Review. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.