

Issues in Direct Writing
- Instruction versus Accountability

Liru Zhang and Robin Taylor
Delaware Department of Education

Background of the Study

- Over the past decades, direct writing has been increasingly used to evaluate students' writing skills in large-scale assessments at the national, state, and local levels. This movement has been paralleled in many classrooms across the country with an emphasis on process approaches in writing instruction.
- Indirect writing, on the other hand, has been criticized lack of validity of measuring student writing skills and having negative impact on writing instruction.

Objectives of the Study

In this study, we will review:

- The Delaware writing assessment
- Technical challenge of direct writing in large-scale assessments
- Role of writing assessment in accountability
- The impact of direct writing on instruction and student achievement

Delaware Writing Assessment (1)

- The DSTP writing assessment is administered to students in grades 3-10.
- The writing assessment includes a standard-alone writing prompt (SAW) and a text-based writing task (TBW).
- A 5-point holistic scoring rubric is used to evaluate student writing skills for both SAW and TBW.

Delaware Writing Assessment (2)

- Student essay is evaluated based on the five important components of writing:
 - Development
 - Organization
 - Sentence structure
 - Words usage
 - Conventions

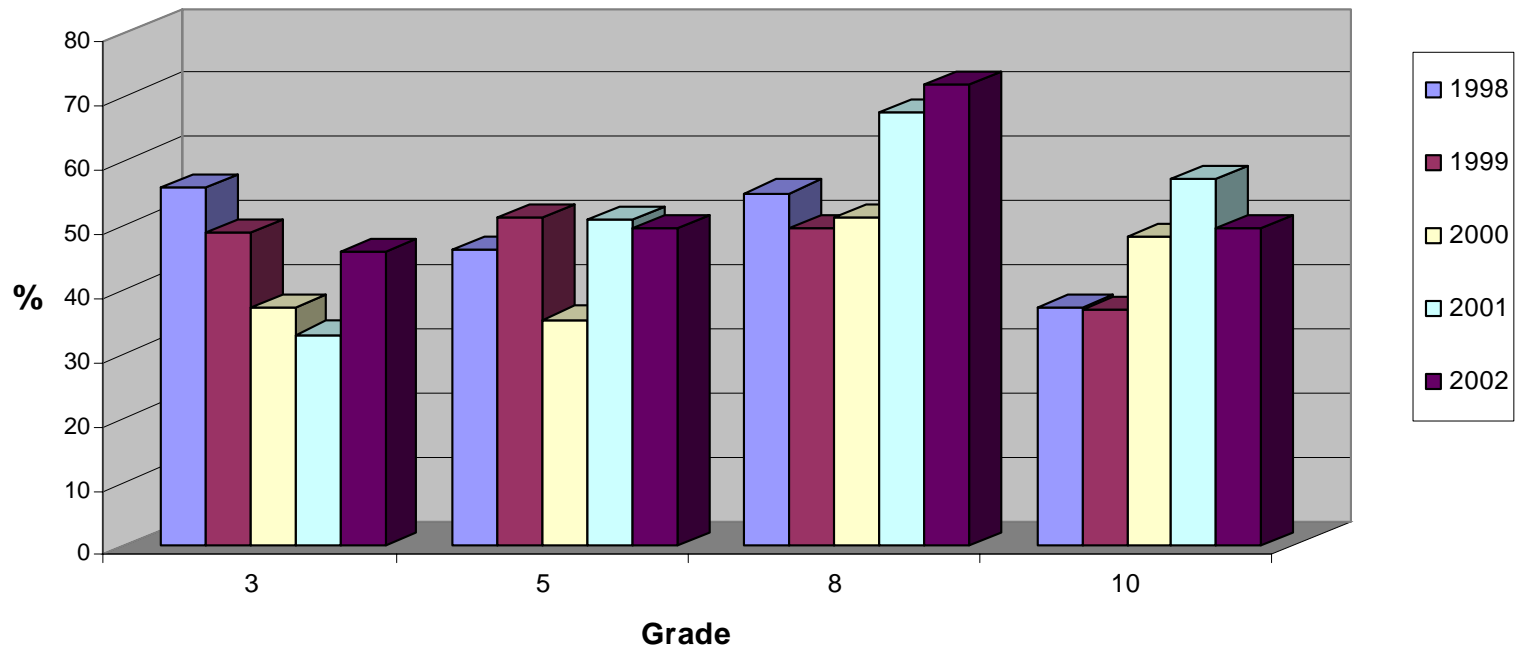
Delaware Writing Assessment (3)

- The writing assessment is designed similar to instructional strategies.
 1. Students are encouraged to use process writing, planning, using a variety of resources (dictionary, check list), drafting, and revising draft.
 2. The writing assessment is not timed.
 3. A total writing raw score, the sum of SAW and TBW scores, is reported.

Challenge of Direct Writing (1)

- Small number of 'items' and relatively narrow scale of test scores of direct writing challenge the traditional reliability.
- The low comparability of writing scores across years and across grades raises concerns of the use of writing scores for high-stakes decisions and accountability system.

Figure 1. Percentage of Meeting the Standard in Writing



Developing Parallel Writing Forms – Option One

Pre-equate the pool of stand-alone writing prompts and use pre-determined criteria for selecting the text-based writing.

Issues:

1. How to report the total writing scores with pre-equated stand-alone writing and the raw scores for text-based writing?
2. How to make the linkage between prompt pools?
3. Will pre-equating reduce or add more errors?

Developing Parallel Writing Forms – Option Two

Use pre-determined criteria to select one discourse of writing prompt for stand-alone and another discourse for text-based writing for 2-3 consecutive years.

Issues:

1. Due to the capacity of the current writing prompt pool, we have to develop a new one to meet the need;
2. Change scoring rubric holistic scoring rubric to specific for each discourse of writing; and
3. Possible negative impact on writing instruction.
4. Cost and time preparing for the new field tests.

Developing Parallel Writing Forms - Option Three

Use indirect MC writing test that focuses on editing and convention to link writing across years.

Issues:

1. The big concern is the validity of indirect MC writing tests. Are they measuring the Delaware Content Standard?
2. Possible negative impact on writing instruction.
3. Does the indirect MC writing tests measure the same construct as direct writing assessment?
4. Additional testing time and cost.

Developing Parallel Writing Forms - Option Four

Use pre-determined criteria to select the text-based and the stand-alone writing. Report the test score on the same 1-15 raw score scale. Stop cross-year comparisons.

Issues:

1. Is it politically possible?
2. How to handle repeated test takers in writing?
3. What is the role of writing score in accountability?
4. How to report writing scores?

Results of Pre-Equating

■ Methodology

1. Two commonly used methodologies on direct writing assessments, Linear Equating and Mean Equating were applied to the 2000 prompt pool for SAW.
2. Sample size is about 1,000 students per prompt.
3. Test score is the sum of two readers' scores on a 0-10 scale.

Table 1a. Results of Mean Equating for Grade 10

Form	F1	F2	F3	F4	F5	F6
N	992	1019	1015	1015	993	1018
Mean	5.524193548	4.747791953	5.132019704	4.963546798	5.294058409	4.85559921
SD	1.623310067	1.374830774	1.690065344	1.797887	1.697998984	1.75818722
Equated	-0.668594334	0.107807261	-0.276420490	-0.107947584	-0.438459195	0
Score	0.331405666	1.107807261	0.723579510	0.892052416	0.561540805	1
	1.331405666	2.107807261	1.723579510	1.892052416	1.561540805	2
	2.331405666	3.107807261	2.723579510	2.892052416	2.561540805	3
	3.331405666	4.107807261	3.723579510	3.892052416	3.561540805	4
	4.331405666	5.107807261	4.723579510	4.892052416	4.561540805	5
	5.331405666	6.107807261	5.723579510	5.892052416	5.561540805	6
	6.331405666	7.107807261	6.723579510	6.892052416	6.561540805	7
	7.331405666	8.107807261	7.723579510	7.892052416	7.561540805	8
	8.331405666	9.107807261	8.723579510	8.892052416	8.561540805	9
	9.331405666	10.107807261	9.723579510	9.892052416	9.561540805	10

Table 1b. Results of Mean Equating for Grade 10

Form	F1	F2	F3	F4	F5	F6
N	992	1019	1015	1015	993	1018
Mean	5.524193548	4.747791953	5.132019704	4.963546798	5.294058409	4.85559921
SD	1.623310067	1.374830774	1.690065344	1.797887	1.697998984	1.75818722
Rounded	-1	0	0	0	0	0
Score	0	1	1	1	1	1
	1	2	2	2	2	2
	2	3	3	3	3	3
	3	4	4	4	4	4
	4	5	5	5	5	5
	5	6	6	6	6	6
	6	7	7	7	7	7
	7	8	8	8	8	8
	8	9	9	9	9	9
	9	10	10	10	10	10

Table 2a. Results of Linear Equating for Grade 10

Form	F1	F2	F3	F4	F5	F6
N	992	1019	1015	1015	993	1018
Mean	5.524193548	4.747791953	5.132019704	4.963546798	5.294058409	4.856
SD	1.623310067	1.374830774	1.690065344	1.797887	1.697998984	1.758
Slope	1.083087733	1.278839006	1.040307245	0.977918646	1.035446569	
Intercept	-1.12758705	-1.21606233	-0.483278063	0.00165425	-0.6261154	
Equated	-1.127587054	-1.216062330	-0.483278063	0.001654250	-0.626115400	0
Score	-0.044499321	0.062776676	0.557029182	0.979572896	0.409331169	1
	1.038588412	1.341615682	1.597336427	1.957491542	1.444777738	2
	2.121676145	2.620454688	2.637643672	2.935410188	2.480224307	3
	3.204763878	3.899293694	3.677950917	3.913328834	3.515670876	4
	4.287851611	5.178132700	4.718258162	4.891247480	4.551117445	5
	5.370939344	6.456971706	5.758565407	5.869166126	5.586564014	6
	6.454027077	7.735810712	6.798872652	6.847084772	6.622010583	7
	7.537114810	9.014649718	7.839179897	7.825003418	7.657457152	8
	8.620202543	10.293488724	8.879487142	8.802922064	8.692903721	9
	9.703290276	11.572327730	9.919794387	9.780840710	9.728350290	10

Table 2b. Results of Linear Equating for Grade 10

Form	F1	F2	F3	F4	F5	F6
N	992	1019	1015	1015	993	1018
Mean	5.524193548	4.747791953	5.132019704	4.963546798	5.294058409	4.856
SD	1.623310067	1.374830774	1.690065344	1.797887	1.697998984	1.758
Slope	1.083087733	1.278839006	1.040307245	0.977918646	1.035446569	
Intercept	-1.12758705	-1.21606233	-0.483278063	0.00165425	-0.6261154	
Rounded	-1	-1	0	0	-1	0
Score	0	0	1	1	0	1
	1	1	2	2	1	2
	2	3	3	3	2	3
	3	4	4	4	4	4
	4	5	5	5	5	5
	5	6	6	6	6	6
	6	8	7	7	7	7
	8	9	8	8	8	8
	9	10	9	9	9	9
	10	12	10	10	10	10

To Equate or Not To Equate?

- There are two reasons for not conducting an equating:
 1. Equating is unnecessary because the forms are 'enough alike' that conducting an equating would likely add in more error than it would remove.
 2. Equating is unwarranted because there would not be significant change between forms.

To Equate or Not To Equate?

- The results from the two equating methods are consistent:
 1. The six forms are similar enough that no significant changes or adjustment could make between forms.
 2. For example in grade 10 Form 1, the rounded scores shifted to the lower end and add a negative score point to the scale.
 3. In most cases, the magnitude of difference was a half or less than a half score point.

To Equate or Not To Equate?

- It has been a challenge to use traditional adjustment of difficulty for direct writing.
 1. The limited number of integer score points does not allow such adjustment.
 2. Convert raw score to a wider scale-score scale (e.g., 0-100) artificially disguise the problem and most scores will not be attainable.
 3. Convert raw score to a scale with decimal point would lack of meaningful interpretation for decimal points and replace the widely-separated integer scores with separate frequency bumps.

Decision - Not Equate

- Decision on the DSTP writing assessment are:
 1. No equating;
 2. Cross year comparison is not encouraged;
 3. Annual report includes descriptive statistics, frequency distributions, and effective size of comparing with previous years scores in addition to the percentage of performance levels;
 4. Release anchor papers and image of all student writing.

Role of Writing Assessment in Accountability (1)

- Student accountability:
 - Grades 3 – 7, and 9; no consequences
 - Grades 8 and 10; a \$1,000 Scholarship for the first 150 highest scoring students in writing each grade
 - Grade 10; writing score is used as part of the ‘Graduation Index’ (20%)

Role of Writing Assessment in Accountability (2)

- School and District Accountability:
 - English language arts percent proficient calculation based on
 - 90% of reading score ($.90 * (\% \text{ of students meeting the reading standard})$)
 - 10% of writing score ($.10 * (\% \text{ of students meeting the writing standard})$)

Impact of Direct Writing (1)

- The Delaware writing assessment has positive impact on classroom instruction in writing:
 - The survey results show that more teachers use process writing in classroom instruction and encourage students using process writing.
 - For example in grade 5, the percentage of students reported that they always use process writing increased from 42% in 1999 to 59% in 2003.

Table 3. The Percentage of Students Using Process Writing

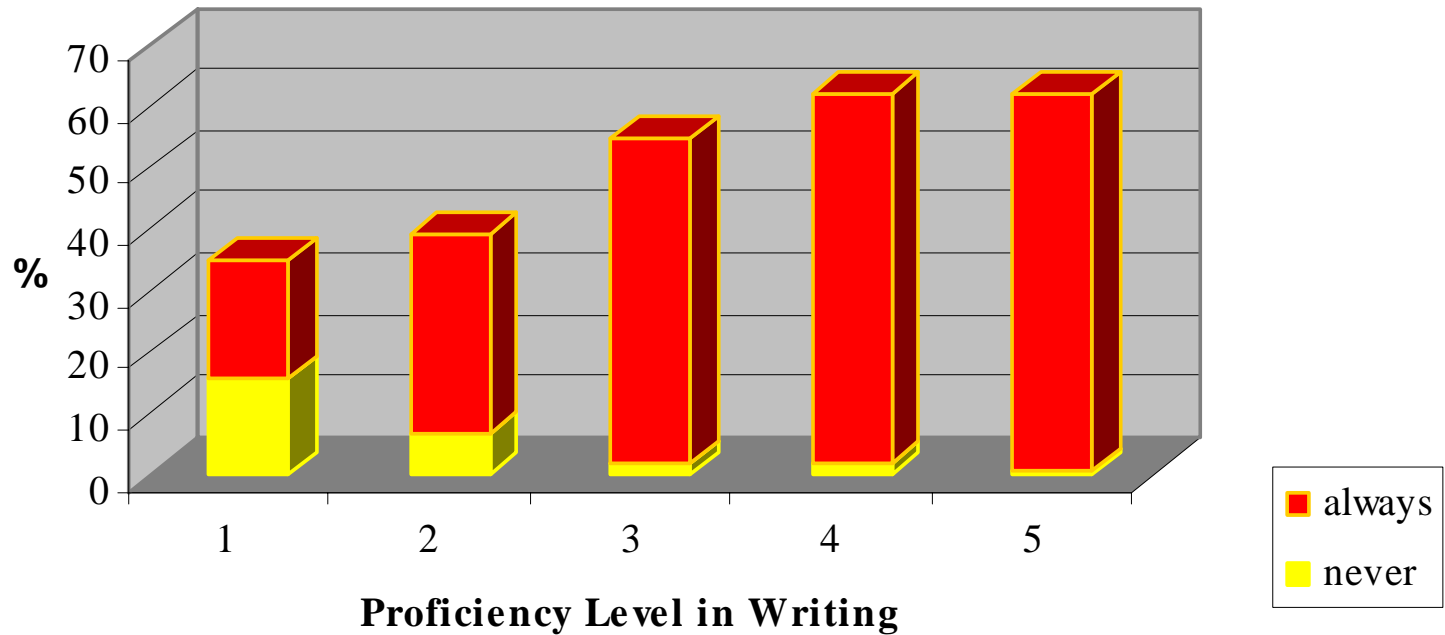
	1999		2000		2002		2003	
GR	Always	ST.	Always	ST.	Always	ST.	Always	ST.
3	38	37	46	34	55	30	55	30
5	42	33	48	31	58	29	59	29
8	39	33	46	31	56	28	55	28
10	40	31	44	31	50	29	48	31

Impact of Direct Writing (2)

- A consistent, positive relationship was found across grades between students using process writing and their performance in writing.

For example, in grade 10, 61% of students in level 5, 60% level 4, 53% level 3, 32% level 2, and 19% in level 1 reported that they always use process writing; in contrast, 16% of students in level 1, 7% level 2, 2% level 3, 2% level 4, and 1% level 5 reported that they never or hardly ever used process writing techniques.

Figure 2. Relationship between Using Process Writing and Student Performance in Writing for Grade 10



Impact of Direct Writing (3)

- Delaware student performance on the 2002 NAEP writing assessment show great improvement.
 - The average score is 163 for grade 4, which is 10 score points higher than the national average (153).
 - The average score is 159 for grade 8, which is 7 score points higher than the national average (152) and increased by 15 points since 1998.
 - The average writing score for grade 8 shows a statistically significant improvement at the .05 level.