Inclusive Practices Reflection Tool: Sample Analysis and Recommendations

Overview

The Inclusive Practices Reflection Tool (IRPT) assesses schools' and school districts' LRE/EE support. The IRPT consists of 34-47 items that rate specific school constructs from "in place" to "not in place" on a four-point lykert scale. In addition, a second three-point lykert scale construct assesses a rater's priority for program improvement from "high" to "low." Thus two separate question constructs comprise this tool: one assesses knowledge and the other opinion.

IRPT administrators collected three separate samples from local schools. Survey raters included general educators, special education teachers, administrators, and support staff. The three samples included: a) forty-percent of staff and faculty from Rockford East High School, b) a small random sample from Guilford High School, and c) a consensus group from Jefferson High School.

In addition, one expert rater independently assessed each school. The same expert rater assessed all three schools using the IRPT.

The following report analyzes the data collected and provides further school sampling recommendations.

Data Analysis

This author calculated two separate statistics that utilized the expert's ratings. The first statistic assessed the overall item agreement (% Agree) between individual raters and the expert rater. The second statistic this author and an outside contributor (MJ) created[1]. The standard difference (St. Dif) statistic indicated the group's standard deviation from the expert rater.

The charts below indicate the agreement percentage and standard difference.

Current Status Questions

		1-24		25-30		31-32		33-34		Overall	
	N	% Agree	St Dif	% Agree	St Dif	% Agree	St Dif	% Agree	St Dif	% Agree	St Dif
Consensus	1	54.2	NA[2]	100	NA	100	NA	50	NA	64	NA
Random	7	31	1.14	52	.680	50	1.30	36	.996	37	1.06
Forty %	70	34	1.31	48	.977	46	1.34	33	1.46	37	1.26

Priority for Improvement

		1-24		25-30		31-32		33-34		Overall	
	N	% Agree	St Dif	% Agree	St Dif	% Agree	St Dif	% Agree	St Dif	% Agree	St Dif
Consensus	1	46	NA	50	NA	50	NA	0	NA	40	NA
Random	7	35	1.14	31	.964	41	1.16	48	.773	35	1.09
Forty %	70	42	.892	42	.831	45	.941	47	.732	42	.875

All standard difference statistics indicate a significant result ($p < .05$).

[1] This author wrote "created" because I have yet to find a reference for this statistic. I remain skeptical that we actually invented a statistic. Nevertheless, I have yet to see anyone use it.

[2] Standard difference could not be calculated here because only one rating was reported. Standard difference reports a group's variance (i.e. more than one rating) from an expert. See Appendix A.

Statistic Explanation. It is important to understand the rational that this author utilized when deciding upon which statistics to report.

Although this author reported percent agreement, its outcome should be perceived with caution. For example, one may observe that the percent agreement remains relatively high within the consensus group. However, this may partially be due to the low sample size (N=1) rather than inflated agreement. Since the consensus group's Priority for Improvement construct remains comparable to the other two groups, little attention should be paid to the few questions the consensus group scored accurately.

In addition, it is important to understand how to interpret the standard difference (St. Dif) score. As explained earlier, standard difference assesses the variance (or how much difference) the groups' scores are from the expert rater. For example, the lower the standard difference, the less total variation from the expert rater. In addition, this statistic provides an observation of variance between constructs and between groups. Therefore it is important to compare on the scores between groups rather than solely the score.

Analysis. A few issues conceptualized from this report:

- The standard difference remained consistent across both question constructs for the random sample.
- No set of questions vastly outperform another set. All questions' agreement and standard difference remained relatively constant between sampling techniques and question constructs.
- This author can conclude that none of the sampling techniques appeared more consistent compared to the other two. All samples provided moderate agreement and variance compared to the expert rater.

Sampling Technique Analysis

Sample demographic analysis ensures an accurate population observation. As such, the chart below indicates the total and percentage of general educators, special educators, support staff, and administrators per each sample.

	Total	General Educators		Special Educators		Support/Others		Administrators	
	N	N	%	N	%	N	%	N	%
Consensus	6	2	33	2	33	0	0	2	33
Random	7	2	29	1	14	2	29	2	29
Forty %	70	51	73	14	20	5	7	0	0

Post hoc analysis revealed that sampling inconsistencies between groups abound. Not surprisingly, seventy-three percent of the largest sample derived from general educators. Since general educators constitute the overwhelming staff majority at most high schools, they rated the IRPT more often. However this overrepresentation represents a large sampling bias for this group.

Overrepresentation transpired within the consensus group and random sample as well. While administrators and support staff represent key school stakeholders, they

do not constitute thirty percent of the school's population. This overrepresentation again represents a potential sampling bias.

Under-representation also occurred. The clear occurrence appeared in the consensus group where no one represented the support staff. In addition, only one special educator represented that group in the random sample.

Finally, each of the sampling methodologies contains additional problems; however the consensus method presents the largest problem. Consensus sampling engenders a non-independent sample. Since administrators were present, this group could potentially represent a power bias. Educators may have feared to present their opinion because of an anonymity absence. Consequently this method remains potentially less representative.

Recommendation

This author recommends stratified random sampling of thirty individuals based on the data collected. None of the sampling methods revealed better than average accuracy and each contained sampling biases. Therefore survey administration methodology should represent a hybrid of the three previous techniques.

The chart below represents this author's school sampling recommendation.

	Total	General Educators		Special Educators		Support/Others		Administrators	
	N	N	%	N	%	N	%	N	%
Proposed Stratified Sample	*~30*	*10-12*	*~40*	*9-11*	*~30*	*1-3*	*~10*	*3-5*	*~20*

This author's reasons for stratified sampling:

- The small random sample produced consistent data. All of the groups were represented in that sample.
- Stratifying the sample thus ensures that all groups will be represented.
- The total N (30) this author recommends represents a number that many statisticians believe the minimum sample size for accurate data collection. Thus, a small random sample could potentially provide inaccurate data. However, because each group potentially will be represented, a sample population of forty percent appears unnecessary.
- The consensus group sample indicated that stratifying is possible. Consensus group protocol stated that all stakeholder groups should be represented. Since it appeared viable previously, this author assumes that this technique will be possible again.
- Participant selection **must** remain randomized within each group. It is important to remember that administrators will not select individuals to participate. Rather, administrators assign individuals to their respective groups and then randomly select individuals from the group to participate.

Conclusion

Data and sampling analyses revealed an inaccuracy between and within groups. Neither the sampling technique nor question constructs indicated better than average outcomes. Therefore this author recommends that future school samples utilize stratified random sampling. This procedure will ensure a representative sample and produce greater expert rater agreement.

Joshua R. Polanin
Loyola University Chicago
School of Education

Apendix A

The chart below indicates the average variance (or the average distance of the group's rating from the expert rater). This statistic should not be considered equitable to the standard difference statistic. Therefore, caution should be implemented when interpreting this statistic.

Average Variance					
	1-24	**25-30**	**31-32**	**33-34**	**Overall**
CS Questions	.52	0	0	.5	.38
PI Questions	.57	.57	.5	1.5	.62