

An Examination of the Impact of 0-100% Mixed Responding on MMPI-2-RF Content-Based Validity Scales

Rosemary Gutierrez, Jennifer Gilmour, Danielle Burchett

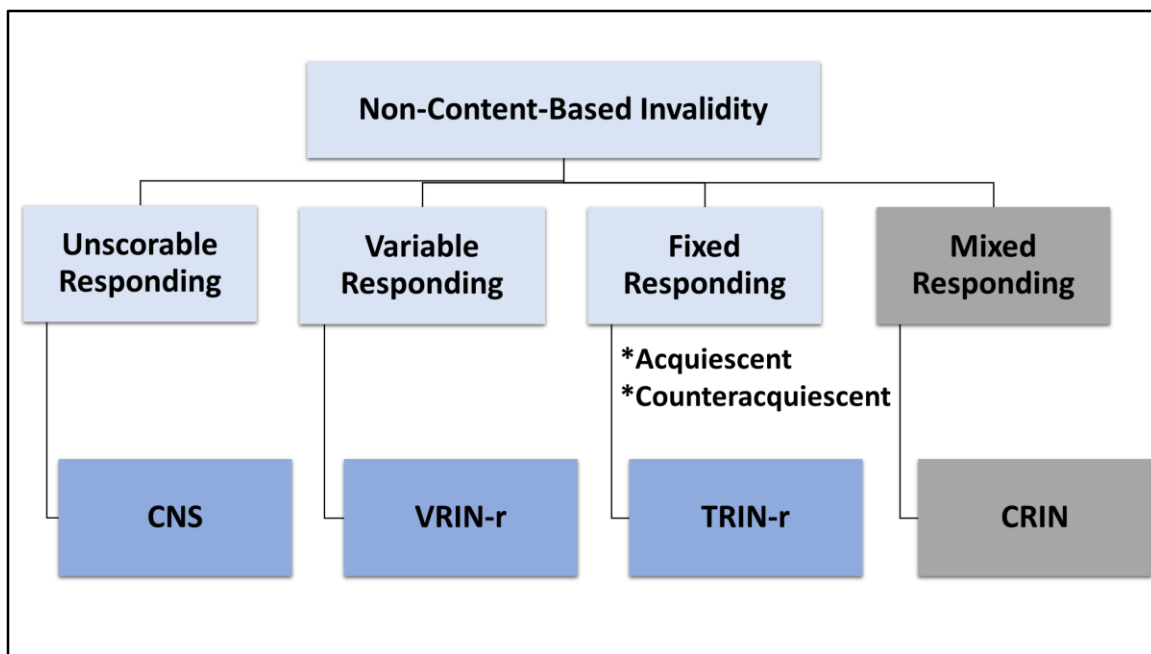
California State University, Monterey Bay

David M. Glassmire

Patton State Hospital



Hello, and thanks for sticking around! My name is Rosemary Gutierrez and I'm an undergraduate at California State University, Monterey Bay. I'll be presenting along with my colleague, Jenny Gilmour.

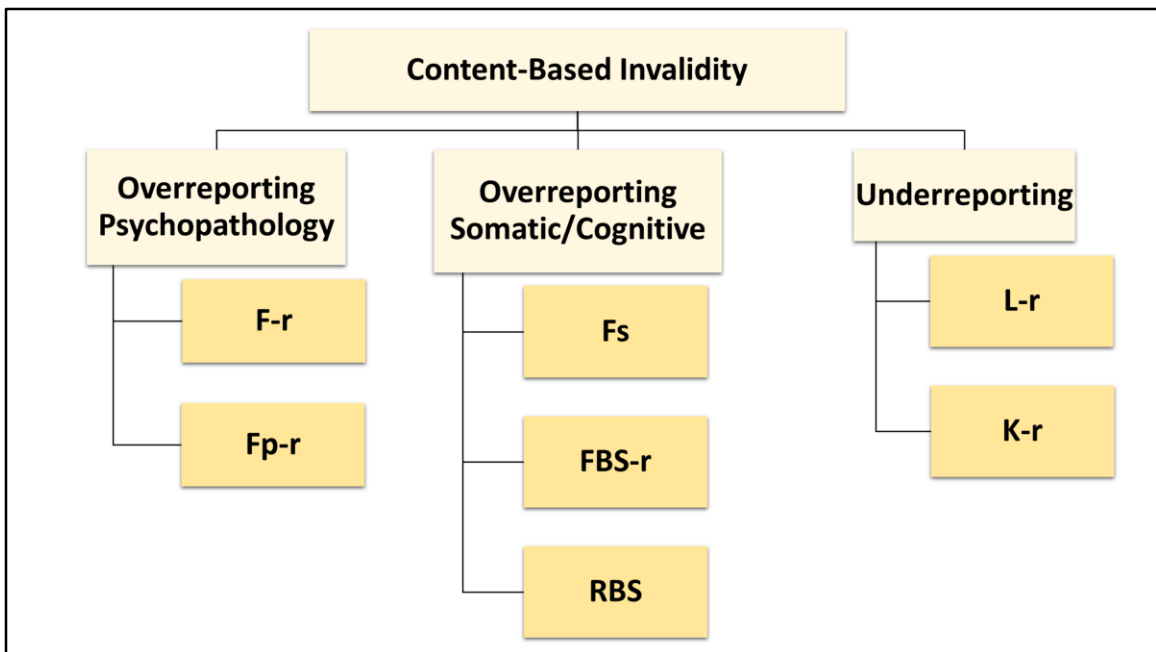


To understand our study, it's important to give a quick reminder about a few threats to the validity of an individual's test protocol, including non-content-based and content-based invalid responding.

Non-content-based invalid responding refers to invalid response styles where people aren't responding to the content of the items. We'll focus on variable responding (which is just marking items at random), and fixed responding (either endorsing a bunch of items as true -- which is acquiescent responding -- or marking a bunch as false -- which is counter-acquiescent responding).

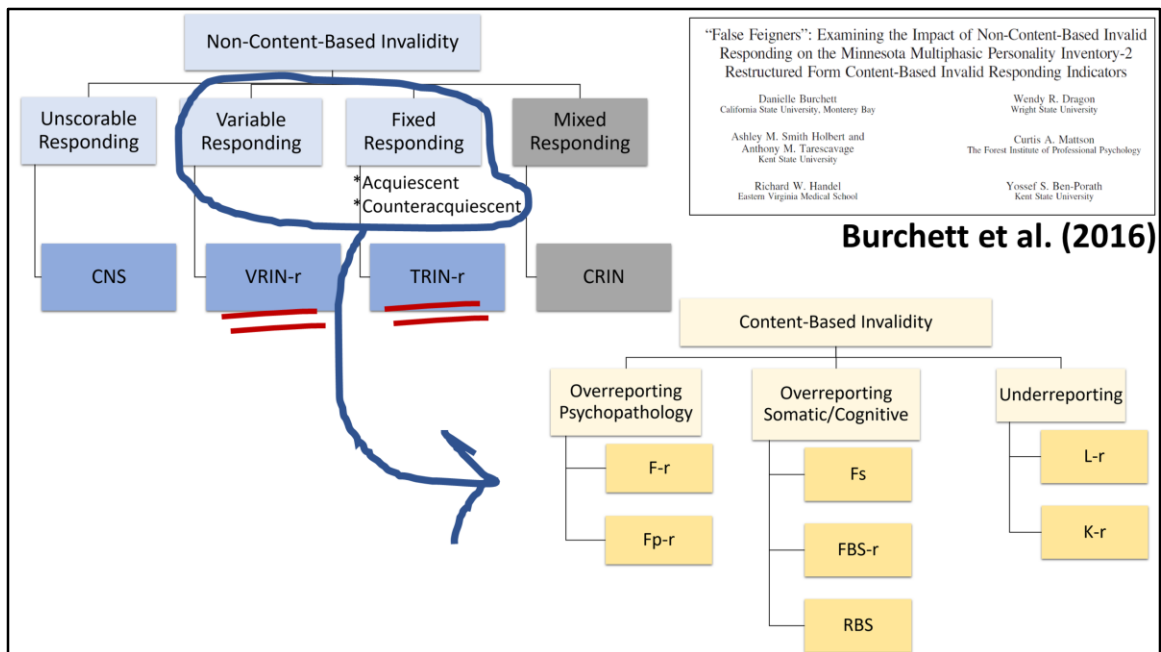
And as we'll discuss today, there's also the possibility of mixed responding, which we define in our study as a mixture of both variable and fixed responding.

You'll note that the MMPI-2-RF includes published indicators for each of these types of invalid responding, with the exception of mixed. We'll talk about an experimental indicator, CRIN, designed for the detection of mixed variable and fixed responding.



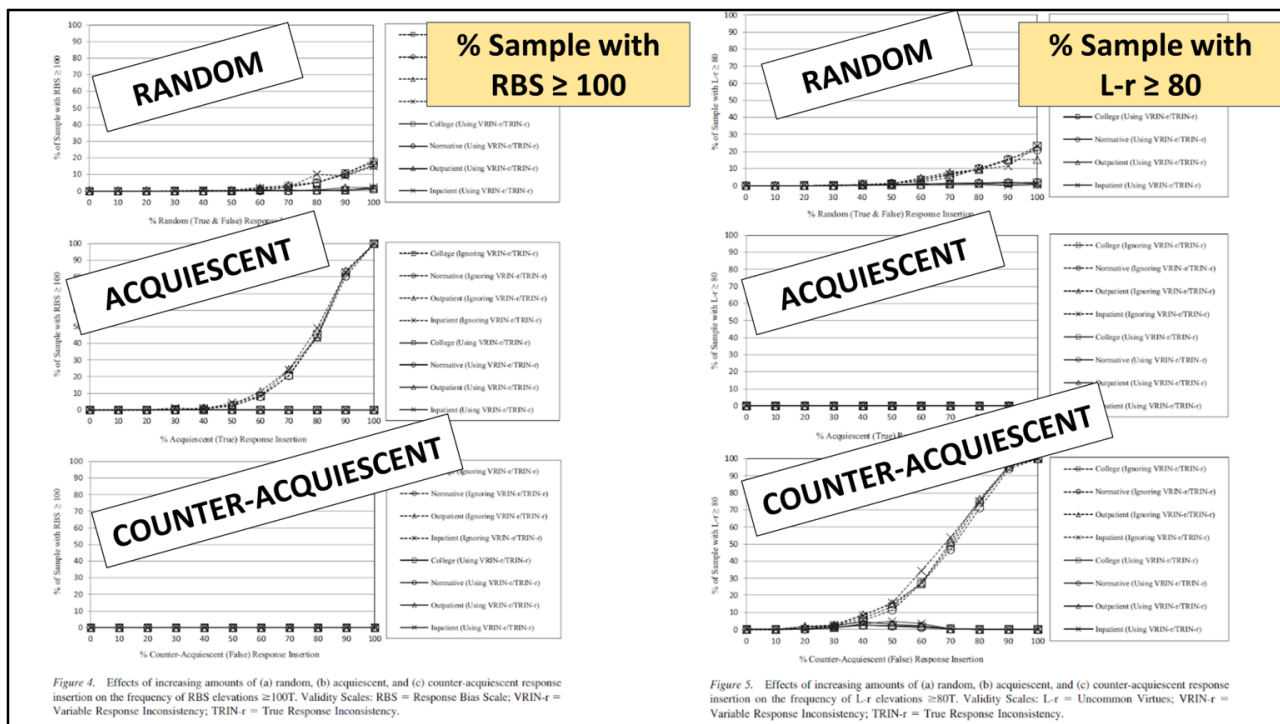
Content-based invalid responding refers to invalid response styles where people ARE responding to the content of the items, but in an invalid manner. They may be overreporting symptoms of psychopathology, somatic, or cognitive problems. Or, they may be underreporting symptoms.

And, as you know, the MMPI-2-RF includes several overreporting and underreporting validity scales.



So, to go into some previous research....Burchett and colleagues wanted to know whether variable, acquiescent, and counter-acquiescent responding would have an impact on the interpretation of the content-based validity scales, both when IGNORING VRIN-r and TRIN-r, and when USING VRIN-r and TRIN-r.

They coined the term “false feigners” to refer to protocols where an individual engages in non-content-based invalid responding, but isn’t detected by VRIN-r or TRIN-r, and their content-based validity scale score elevations incorrectly flag them as an overreporter or underreporter.



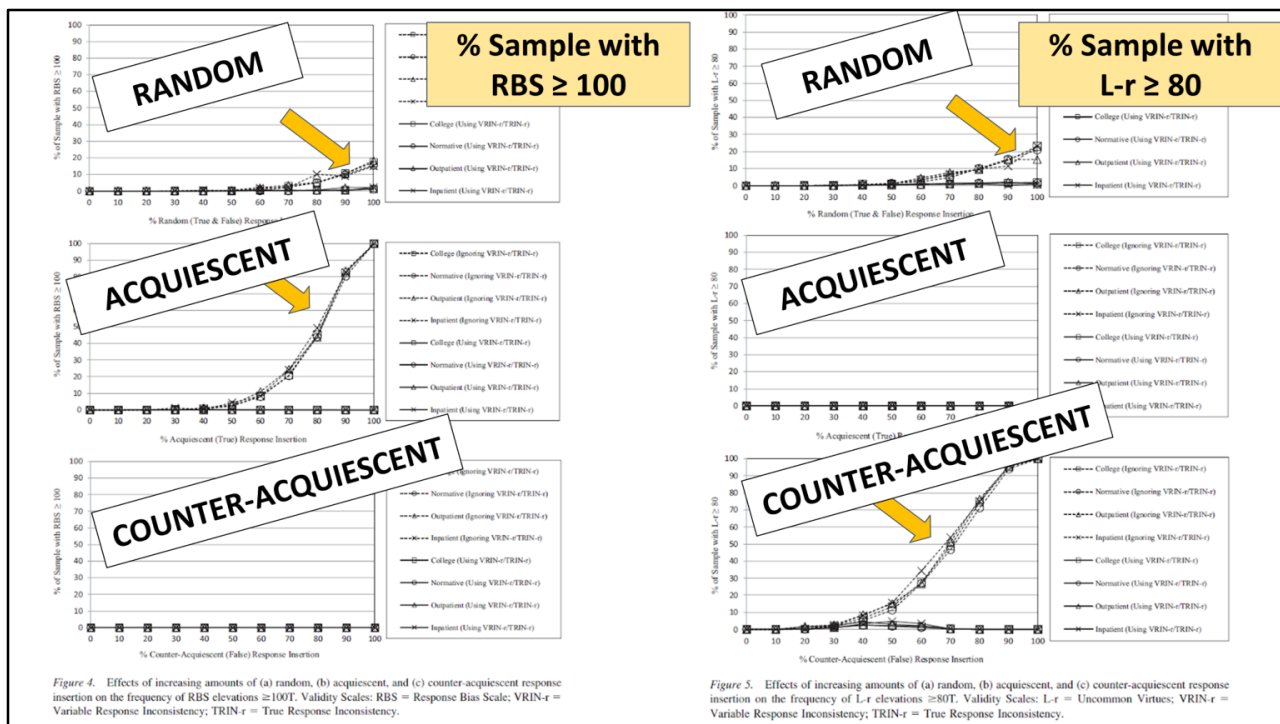
Burchett's 2016 study taught us that....

CLICK

...even though several MMPI-2-RF scales are sensitive to random, acquiescent, and counter-acquiescent responding (as seen with ORANGE ARROWS)...

CLICK

...for most overreporting and underreporting scales, they were either not elevated in the presence of variable and fixed responding OR VRIN-r and TRIN-r were useful in detecting the invalid protocols and, thus, in preventing incorrect "false feigner" interpretations (as seen with GREEN ARROWS).



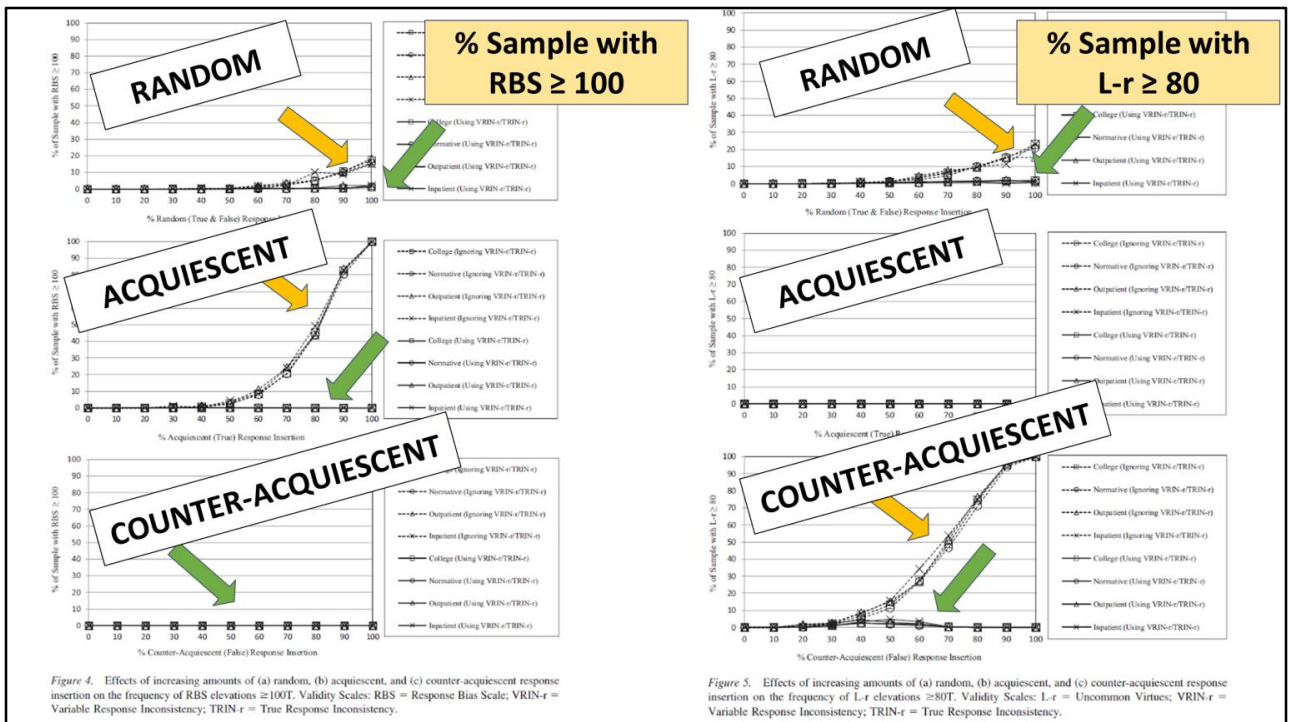
Burchett's 2016 study taught us that....

CLICK

...even though several MMPI-2-RF scales are sensitive to random, acquiescent, and counter-acquiescent responding (as seen with ORANGE ARROWS)...

CLICK

...for most overreporting and underreporting scales, they were either not elevated in the presence of variable and fixed responding OR VRIN-r and TRIN-r were useful in detecting the invalid protocols and, thus, in preventing incorrect "false feigner" interpretations (as seen with GREEN ARROWS).



Burchett's 2016 study taught us that....

CLICK

...even though several MMPI-2-RF scales are sensitive to random, acquiescent, and counter-acquiescent responding (as seen with ORANGE ARROWS)...

CLICK

...for most overreporting and underreporting scales, they were either not elevated in the presence of variable and fixed responding OR VRIN-r and TRIN-r were useful in detecting the invalid protocols and, thus, in preventing incorrect "false feigner" interpretations (as seen with GREEN ARROWS).

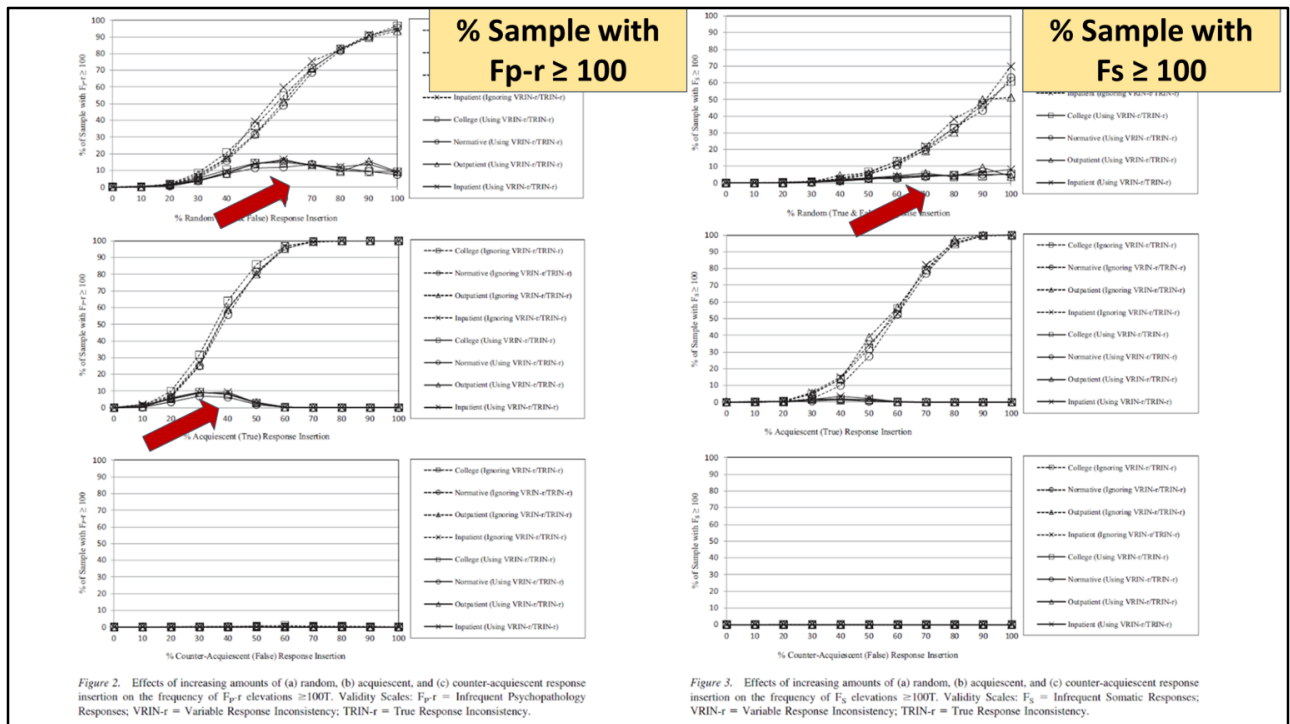


Figure 2. Effects of increasing amounts of (a) random, (b) acquiescent, and (c) counter-acquiescent response insertion on the frequency of F_p-r elevations ≥ 100 . Validity Scales: F_p-r = Infrequent Psychopathology Responses; VRIN-r = Variable Response Inconsistency; TRIN-r = True Response Inconsistency.

Figure 3. Effects of increasing amounts of (a) random, (b) acquiescent, and (c) counter-acquiescent response insertion on the frequency of F_s elevations ≥ 100 . Validity Scales: F_s = Infrequent Somatic Responses; VRIN-r = Variable Response Inconsistency; TRIN-r = True Response Inconsistency.

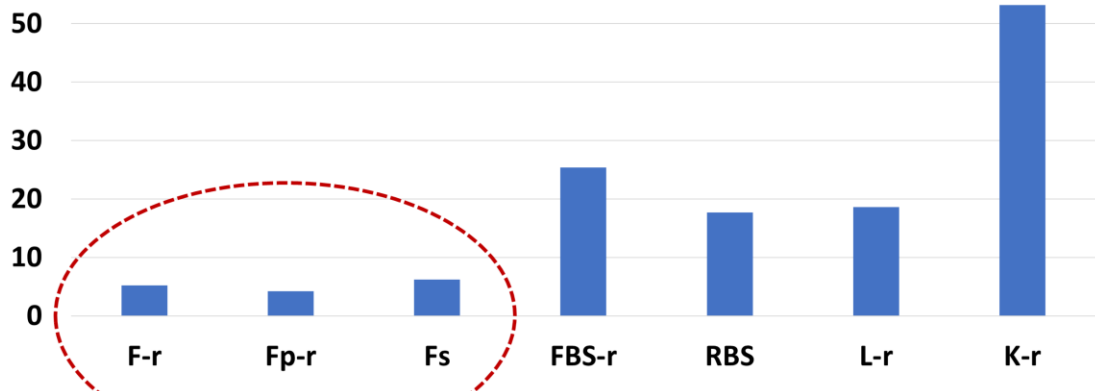
Some minor exceptions include F_p-r and (to a lesser degree) F_s , where, at moderate levels of random and acquiescent responding, some people were NOT flagged by VRIN-r or TRIN-r but their F_p-r or F_s scores were elevated, thus leading to occasional “false feigner” interpretations for these scales.

Importance of Content-Based Validity Scale Characteristics

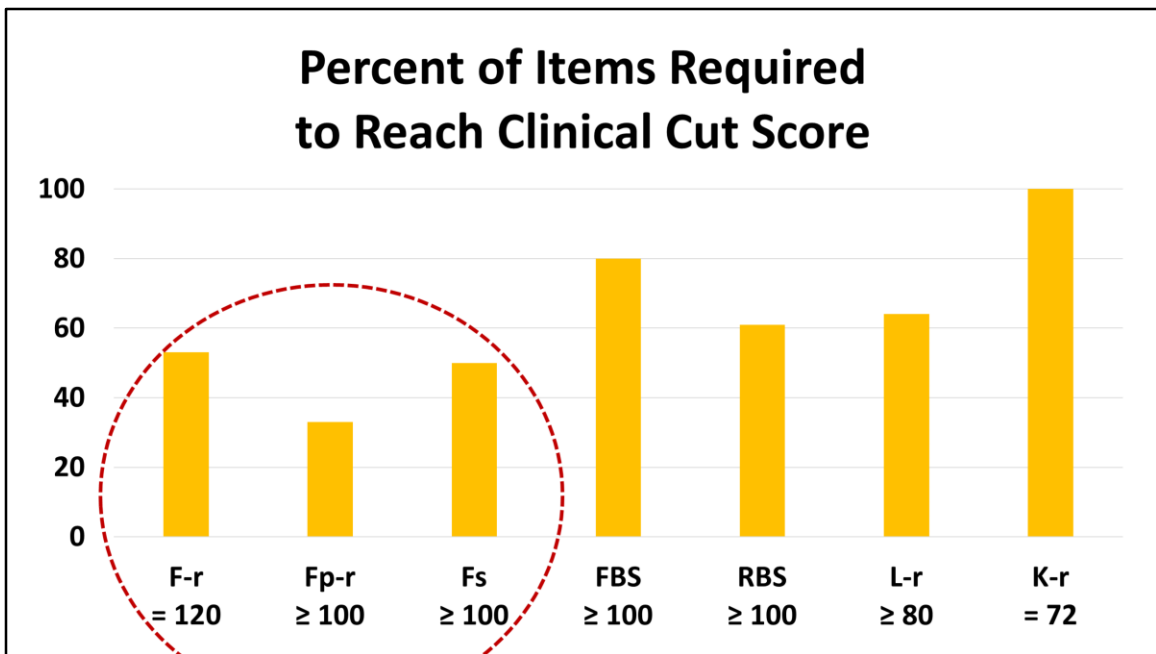
- 1. Item Endorsement Rarity**
- 2. Percentage** of Items Needed to Reach Clinical Cutoffs
- 3. Balance** of True-Keyed and False-Keyed Items

In that study, they noted the links between their findings and several content-based validity scale characteristics.

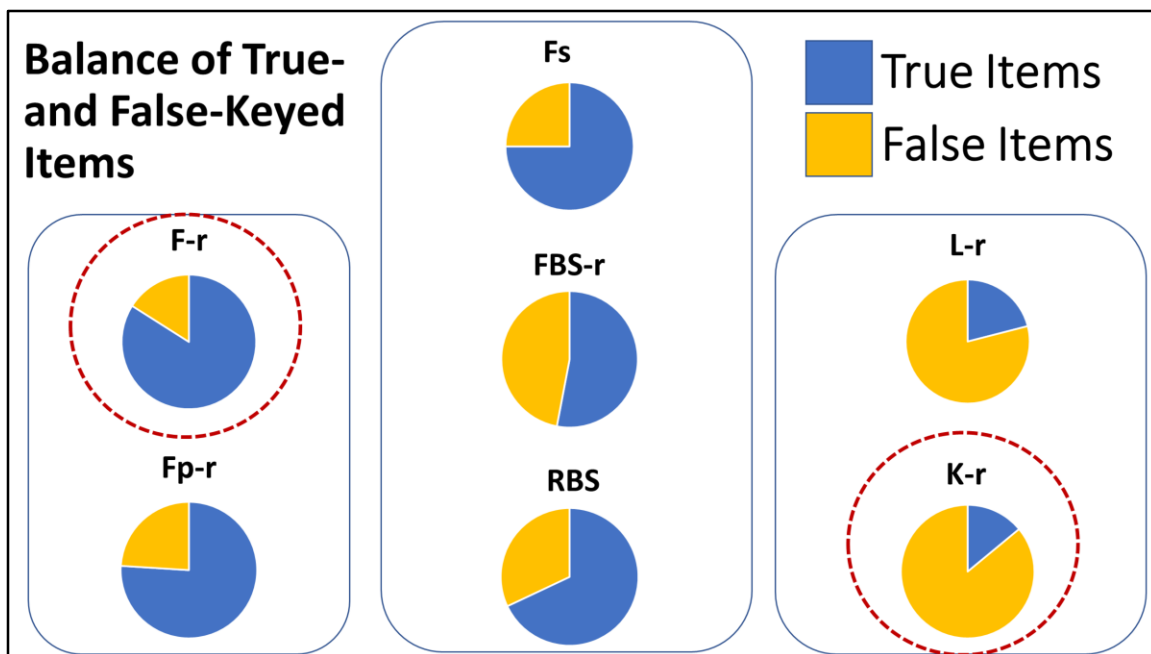
Item Endorsement Rarity (Normative Item Indorsement Index; Burchett et al., 2016)



For example, on average, individual Fp-r, F-r, and Fs items are endorsed much more rarely in the normative sample, compared to items on the other scales. So, while this was by design and makes the scales particularly effective, they may also be especially impacted by random or fixed responding.



They also require a relatively low percentage of their items to be endorsed in order to reach clinical cut scores, so it may take relatively less variable or fixed responding to impact them.



And, when we think about the impact of acquiescent or counter-acquiescent responding, we especially want to pay attention to scales where the balance of True and False items is far from 50% -- such as for F-r (lots of true-keyed items) and K-r (lots of False-keyed items).

Combined Response Inconsistency (CRIN)

- Developed on the MMPI-A-RF (Archer, Handel, Ben-Porath, & Tellegen, 2016) to detect quasi-random & fixed responding
- Whitney et al. (2018a, 2018b) studied an experimental MMPI-2-RF CRIN:



- Using our forensic inpatient sample, it was:
 - sensitive to the impact of mixed responding
 - incrementally useful over VRIN-r and TRIN-r (40% mixed responding)
- Lemaster et al. (2019) found similar results (40-100% mixed responding)

Another important topic to introduce our project is CRIN -- the Combined Response Inconsistency validity indicator, which was developed for the MMPI-A-RF in an effort to flag protocols that had mixed quasi-random and fixed responding.

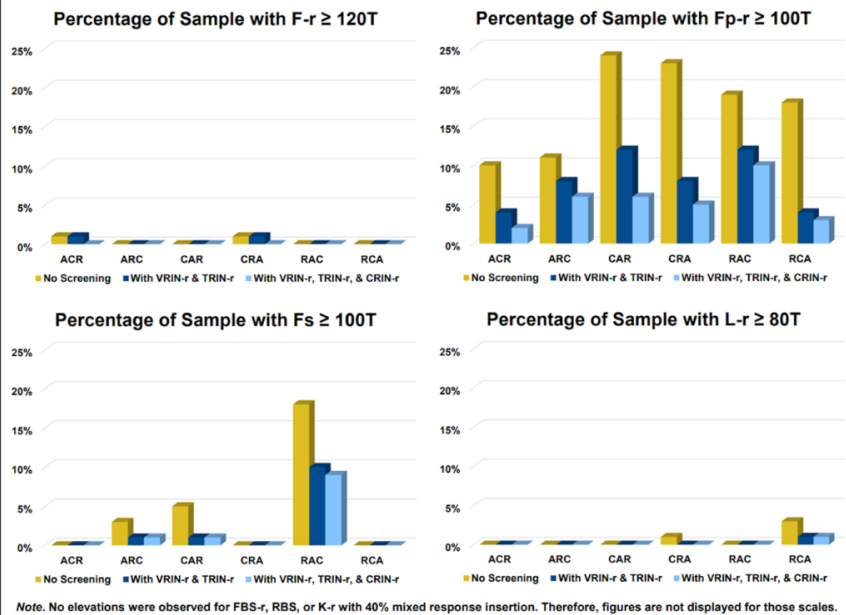
Using an experimental MMPI-2-RF CRIN, Whitney found it was sensitive to mixed responding and incrementally useful over VRIN-r and TRIN-r when 40% of item responses were replaced with computer-generated mixed responding.

Lemaster (2019) replicated the project with college samples and came to the same conclusions, focusing on 40% to 100% mixed responding levels.

Burchett, Gutierrez, Hatch, Chille, & Glassmire (2018)

- Initial exploration of the impact of mixed responding on content-based validity scales (40% mixed responding)
- Fp-r & Fs were most impacted
- VRIN-r and TRIN-r helped substantially
- CRIN added incrementally

Figure 1: MMPI-2-RF Content-Based Validity Scale Clinical Elevation Frequencies Due to 40% Mixed Response Insertion (N = 156)



One last study for us to mention is our MMPI symposium poster from last year, as it was a bit of a pilot study for our current project.

We explored the impact of mixed responding on the content-based validity scales. But, we only focused on data with 40% of items replaced by mixed responding.

Fp-r and Fs were most impacted by mixed responding.

VRIN-r and TRIN-r helped substantially to prevent false feigner rates, but didn't get rid of them completely.

Especially for Fp-r, CRIN was incrementally useful in further reducing false feigner rates.

Aims & Research Questions

- Extension of Burchett et al. (2018) to examine 0% to 100% mixed responding levels



- Like Whitney et al. (2018b), we used 6 mixed responding scenarios



With that study focusing only at 40%, we were interested in expanding the focus on the full spectrum of 0% to 100% computer-generated mixed responding.

Using Whitney's 6 mixed responding scenarios, we were interested in understanding:

Aims & Research Questions

- Extension of Burchett et al. (2018) to examine 0% to 100% mixed responding levels

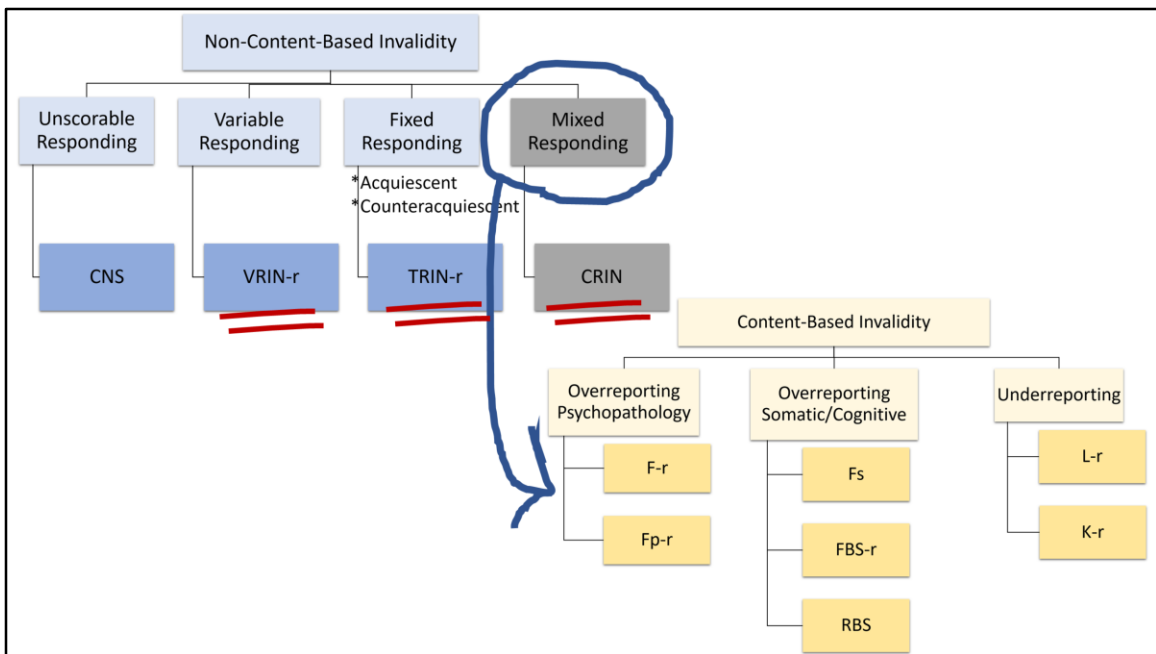


- Like Whitney et al. (2018b), we used 6 mixed responding scenarios



- ***What impact does mixed responding have on content-based validity scales?***
- ***Are VRIN-r & TRIN-r useful in reducing false feigning rates?***
- ***Is CRIN incrementally useful in further reductions?***

- (1) What impact does mixed responding have on content-based validity scales?
- (2) Are VRIN-r and TRIN-r useful in reducing false feigning rates? and,
- (3) Is CRIN incrementally useful in further reductions?



So, we're looking at mixed responding's impact on the content-based scales, first ignoring and then using VRIN, TRIN, and (incrementally) CRIN.

Overall Hypotheses

- Based scale characteristics, mixed responding would impact scale means & sample elevation percentages as follows:

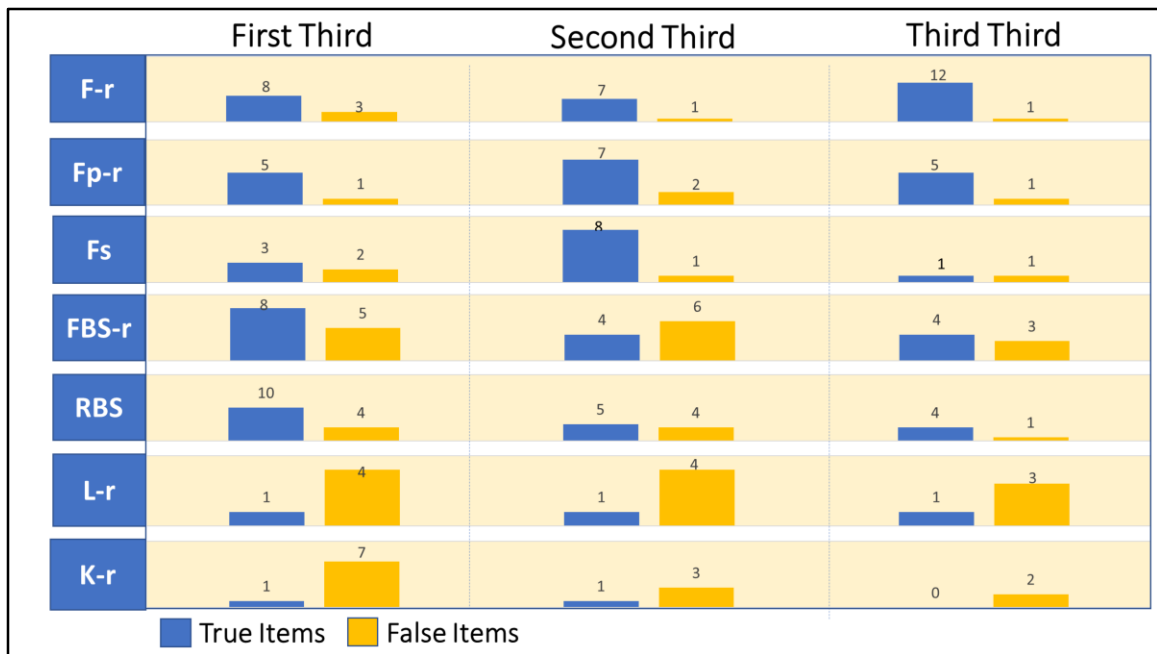
Most Impacted	Moderately Impacted	Least Impacted
Fp-r, F-r, Fs	RBS, FBS-r, L-r	K-r

- VRIN-r & TRIN-r would notably reduce “false feigner” rates
- CRIN would be incrementally useful in further reductions

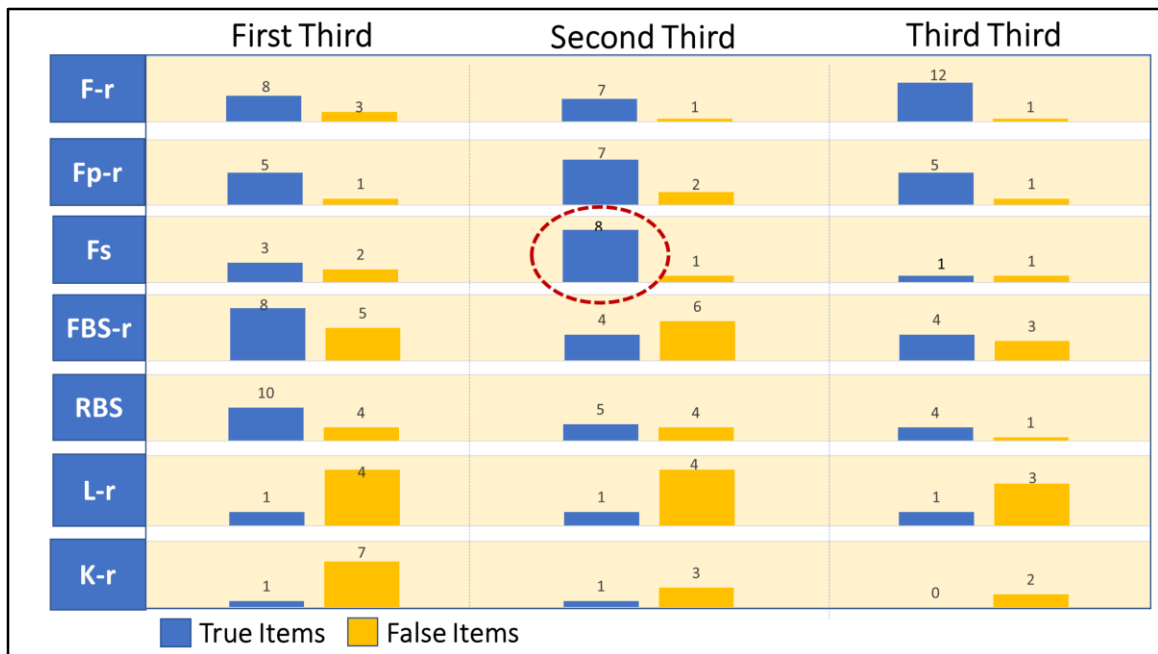
We had three broad hypotheses.

First, after studying the content-based validity scale characteristics, we hypothesized that Fp-r, F-r, and Fs would be most impacted by mixed responding, and that K-r would be least impacted.

Based on previous studies, we hypothesized that VRIN and TRIN would be useful in flagging elevated protocols, and that CRIN would be incrementally useful in flagging even more of them.



After studying the location of content-based validity scale items across three equal sections of the booklet, we also developed some hypotheses about the impact of the six mixed responding conditions, too.



As one example, for Fs, there are quite a few True-keyed items in the middle portion of the test. So, we thought those conditions where acquiescent responding was inserted in the middle portion would be especially impactful for Fs.

Mixed Responding Condition Hypotheses

- Based on **item location** and **% of items needed to elevate**, we developed hypotheses about the impact of the six mixed responding conditions.
- We expected the greatest impacts for...

F-r	CRA, RCA
Fp-r	All Conditions
Fs	RAC, CAR, ARC
FBS-r	ACR, ARC, RCA
RBS	ACR, ARC, RCA
L-r	CRA, RCA
K-r	CAR, CRA

Here is a list of the conditions we expected would have the most dramatic impact on each of the content-based validity scales.

Participants

$N = 1,100$ Forensic Inpatients



Exclude 944 Invalid Protocols
(*strict exclusionary criteria;*
Burchett et al., 2016)

CNS ≥ 15 ; VRIN-r $\geq 70T$, TRIN-r $\geq 70T$, F-r $\geq 79T$, Fp-r $\geq 70T$, Fs $\geq 80T$, FBS-r $\geq 80T$, RBS $\geq 80T$, L-r $\geq 65T$, and/or K-r $\geq 60T$



Final $n = 156$ Forensic Inpatients

We used an archival forensic inpatient dataset. Starting with data from 1,100 patients, we used very strict exclusionary criteria in order to start with a very clean valid dataset before adding in computer-generated mixed responding.

This excluded a very large number of protocols, leaving us with a final sample of 156 inpatients.

Participants

Age:

M (SD) = 42.28 (10.60)

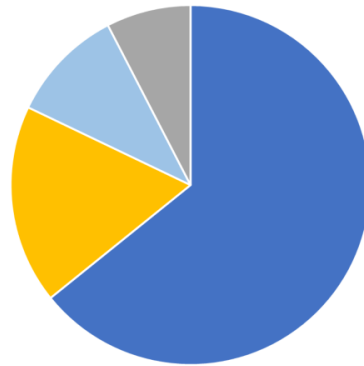


75% Male



25% Female

Ethnicity



■ Caucasian

■ African American

■ Hispanic/Latino

■ Other/Missing

The sample was mostly male, mostly Caucasian or African American, and with a mean age of 42 years.

Measures



- Minnesota Multiphasic Personality Inventory-2 Restructured Form (Ben-Porath & Tellegen, 2008/2011)
 - Published Validity Scales
- Experimental MMPI-2-RF CRIN (Whitney et al., 2018a)



Our study focuses on the published MMPI-2-RF Validity Scales and Whitney's experimental MMPI-2-RF CRIN.

Procedure

- 61 Datasets (0-100% Mixed Response Insertion x 6 Conditions)

	0	10	20	30	40	50	60	70	80	90	100
ACR	BASELINE DATA										
ARC											
CAR											
CRA											
RAC											
RCA											

A: Acquiescent (in First Section)

C: Counteracquiescent (in Second Section)

R: Random (in Third Section)

We used the clean sample and then created 60 simulation datasets with 10% to 100% of items replaced with mixed responses.

We also used 6 mixed responding conditions.

As an example, ACR means we inserted acquiescent responses in the first third of items, counter-acquiescent in the middle third, and random responding in the final third.

Procedure

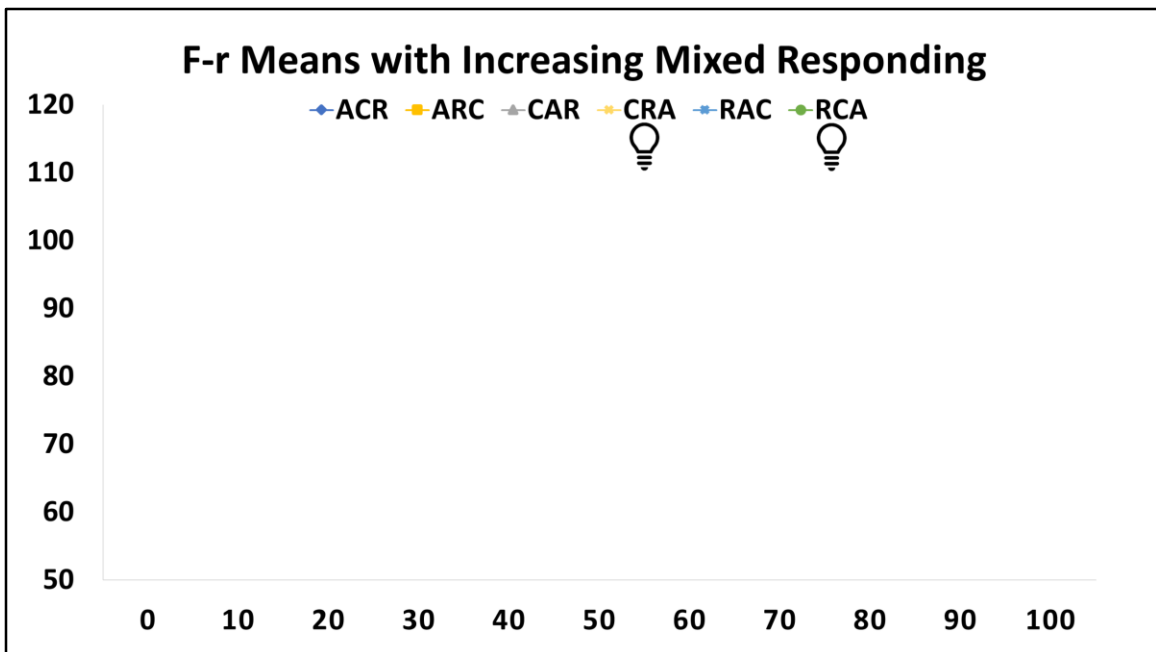
- We Examined:
 - **Mean Scores** for Content-Based Validity Scales
 - **Percentage of Sample** with Scores Above Clinical Cut Score
 - **Overall** (No Screening)
 - After Removing Invalids Using **VRIN-r & TRIN-r**
 - After Removing Invalids Using VRIN-r, TRIN-r, & **CRIN**

[Cut?]

We examined mean scores and the percentage of each sample with scores above designated clinical cut scores (1) without screening using VRIN, TRIN, or CRIN, (2) after using VRIN and TRIN, and (3) after using CRIN.

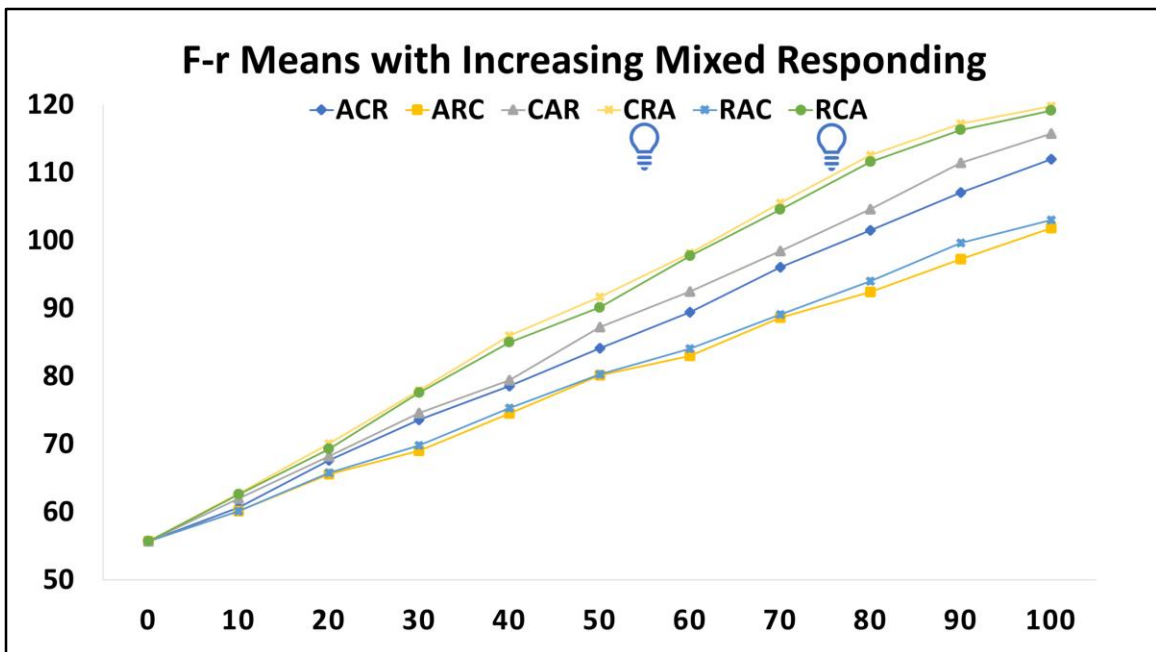
RESULTS

Hi, I'm Jenny. Let's talk about the findings.



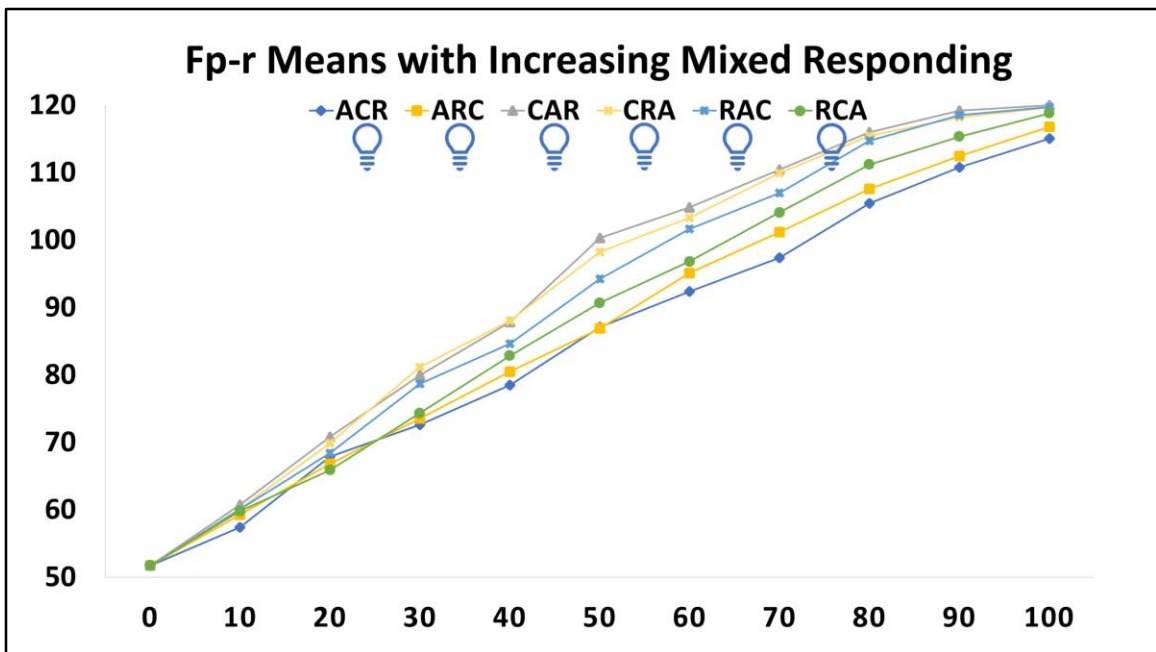
So, first, we'll take a look at those mean scores for each content-based validity scale.

You'll see little light bulbs up here, which remind us of the conditions we thought would be most impacted by mixed responding.

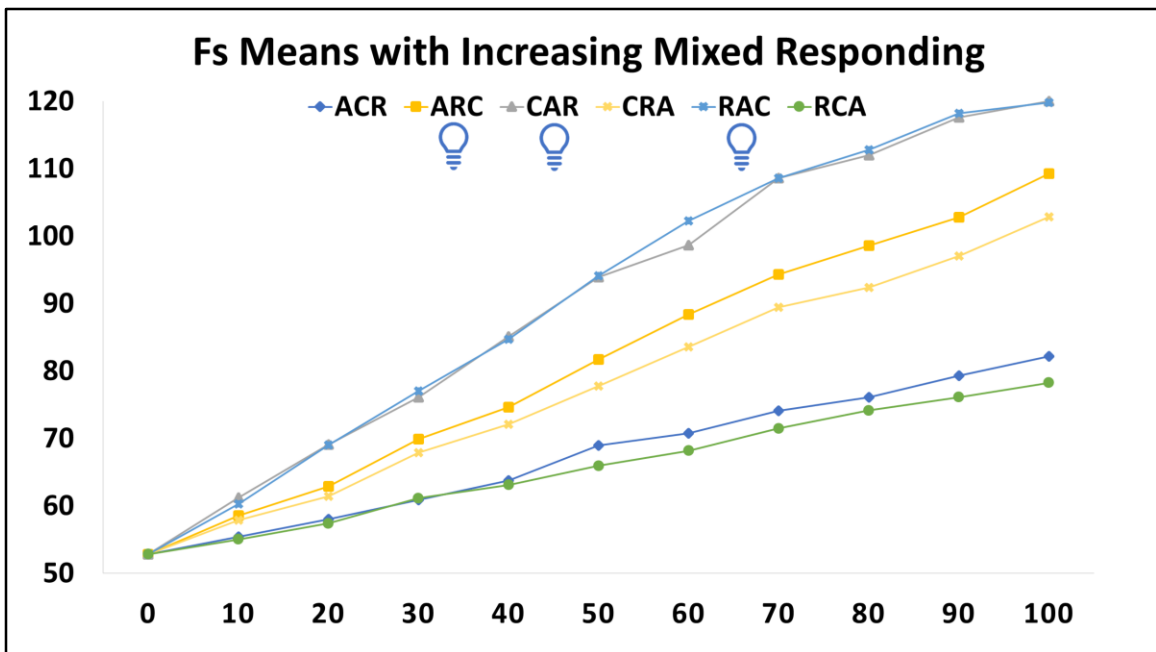


As you'll see for many of the scales, as we added more and more mixed responding, mean scores for F increased, from about 55T up to nearly 120T.

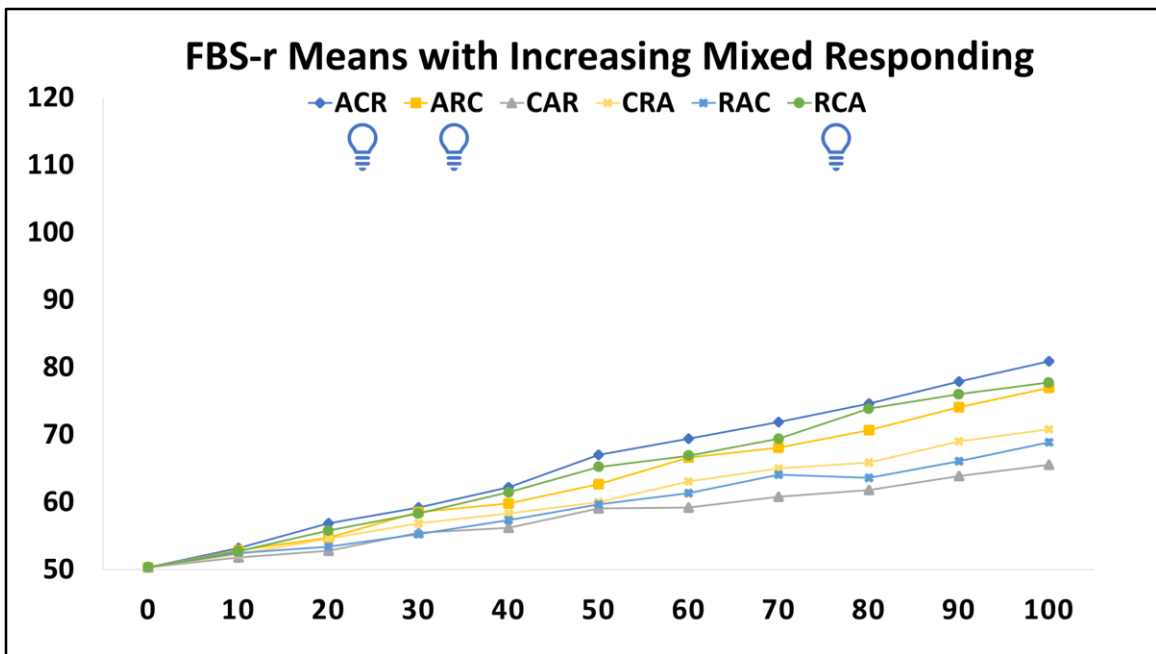
As we hypothesized for F, the conditions in which acquiescent responses were inserted in the last third of the test elevated F mean scores the most.



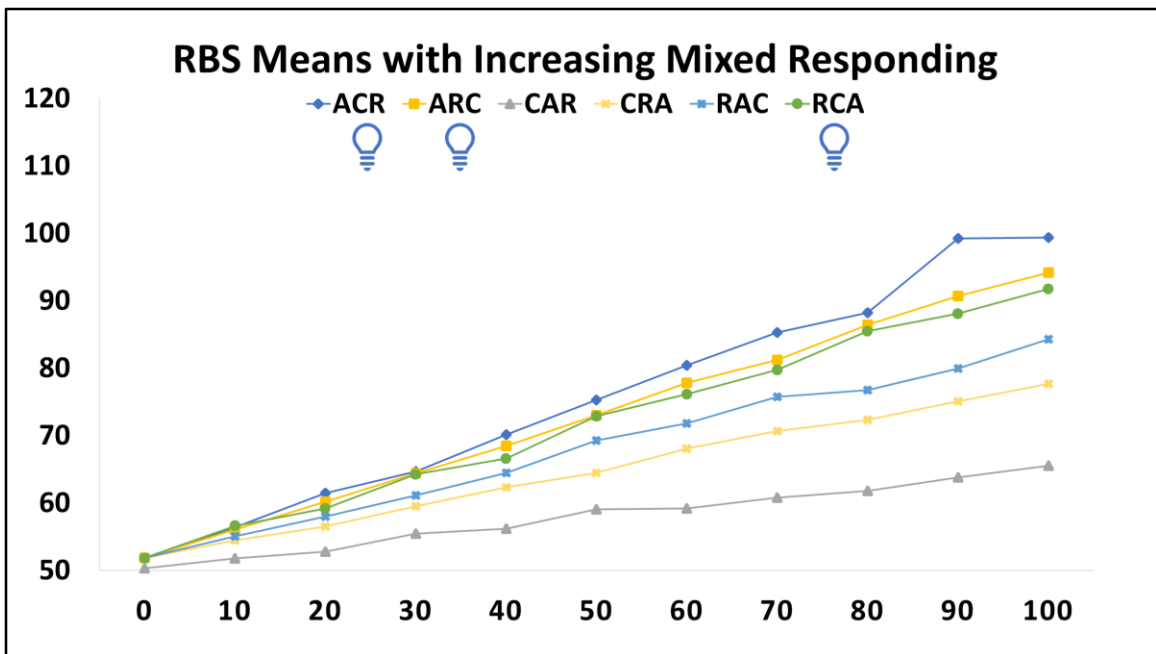
For Fp, we anticipated all conditions would have significant impacts on the scale due to its low item endorsement rate in the normative sample. And, that's what we found.



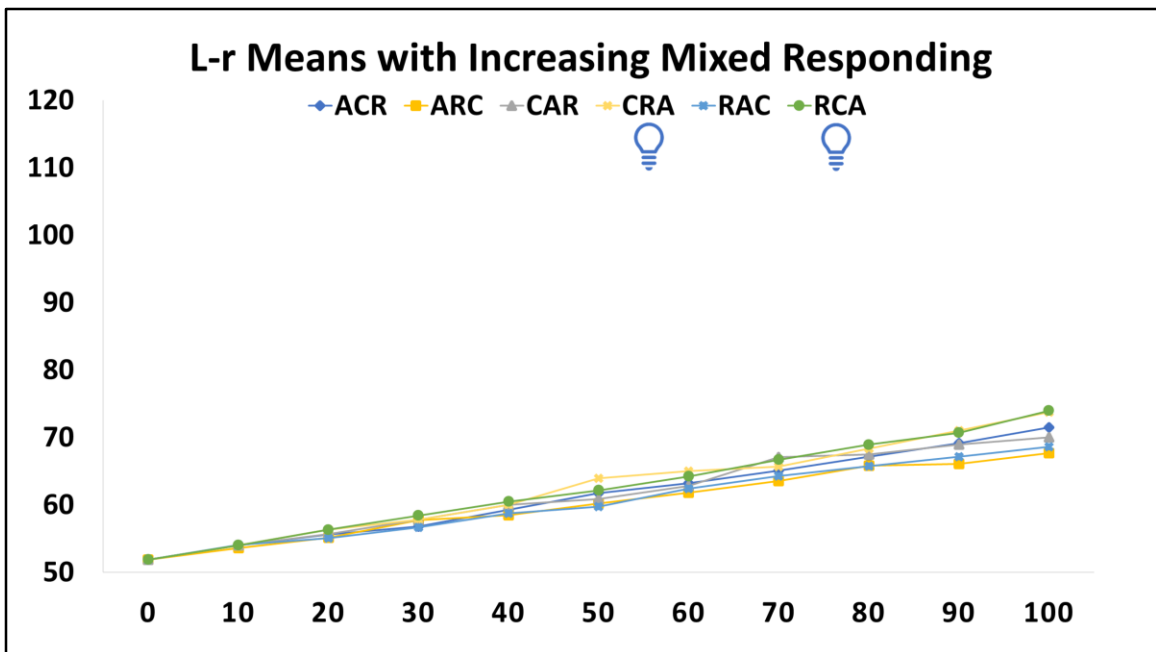
For Fs, we expected ARC, CAR, and RAC would have the highest elevations because of item placement and true/ false ratio of items. That's consistent with what we found.



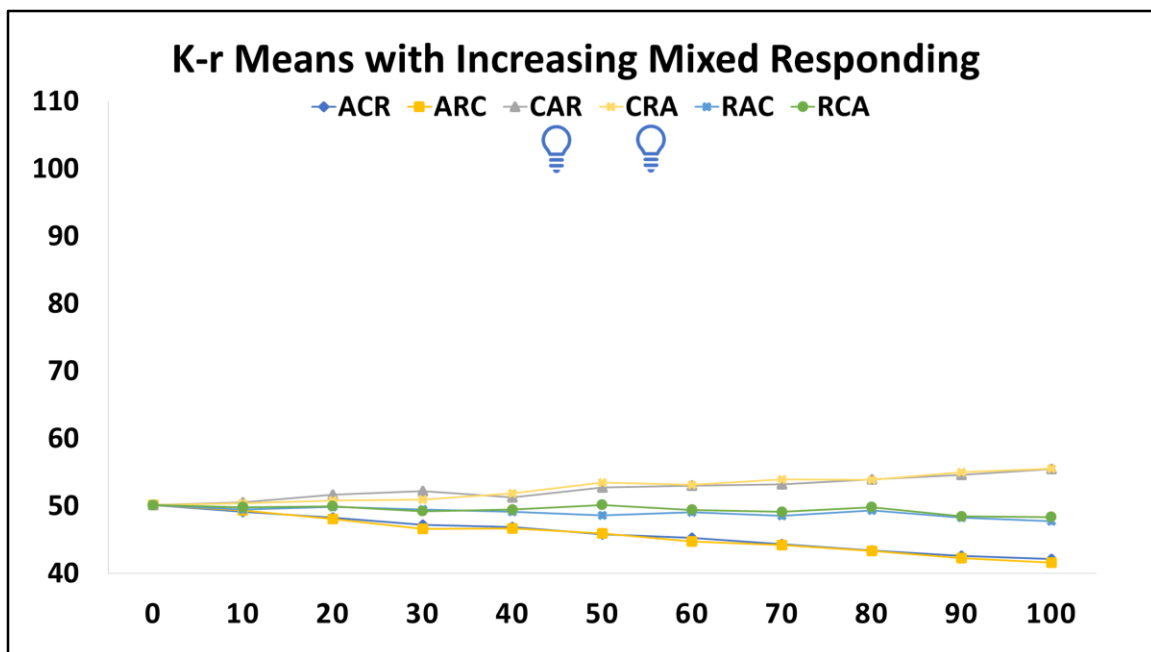
For FBS, we expected conditions ACR, ARC, and RCA to have some impact on the scale, and those were the three conditions with the highest means.



Similar to FBS, we also hypothesized the same conditions would most impact RBS. And indeed, those were the most impacted.



L has more true-keyed items in the last third of the test, so we expected conditions CRA and RCA to be the most impacted. We found that to be true, although there wasn't much variability in L scores across conditions.



Lastly, for K, because there are a lot of false-keyed items on the first third of the booklet, we hypothesized two conditions would have an impact on mean scores: CAR and CRA. And that was the case.

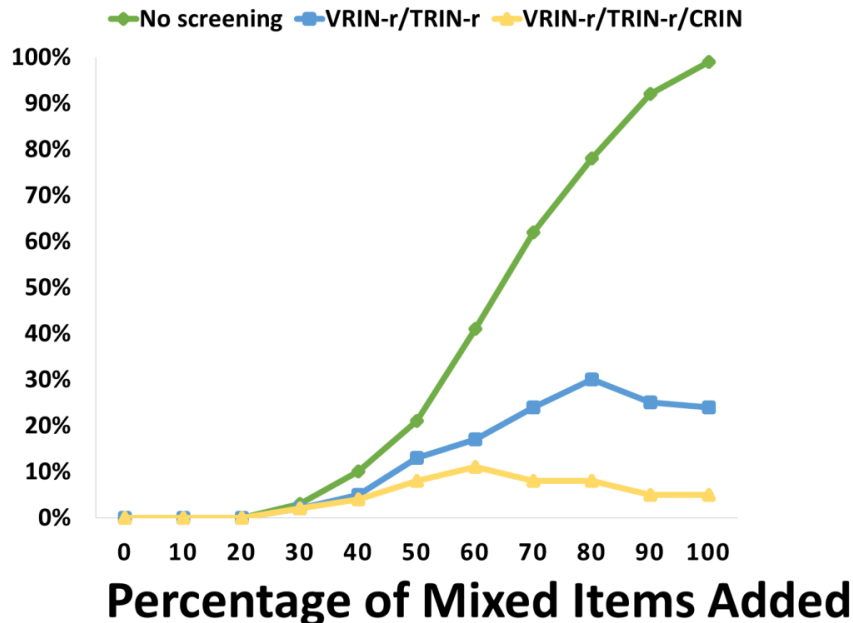
Interestingly, when acquiescent or random responses were placed in the first third of the test, we saw a slight decrease in mean scores.

Be ready to answer question about this – ALSO, K-r is the only scale where NIEI was above 50. We may be dropping endorsement with random/fixed responding.

Fp-r ≥ 120

ARC

**Percentage
of Sample
Elevated to
Clinical Cut
Score**

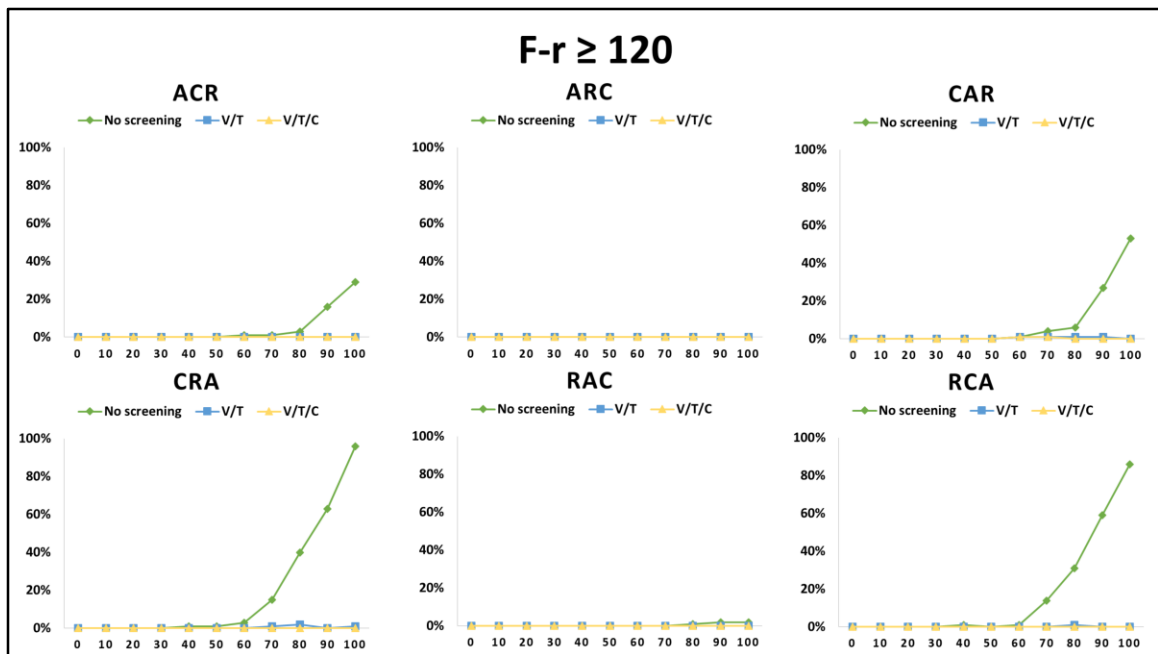


Next, we'll look at the results that show what percentage of the sample reached clinical cutoffs on the content-based validity scales. I'll orient you with an example – Fp in the ARC condition.

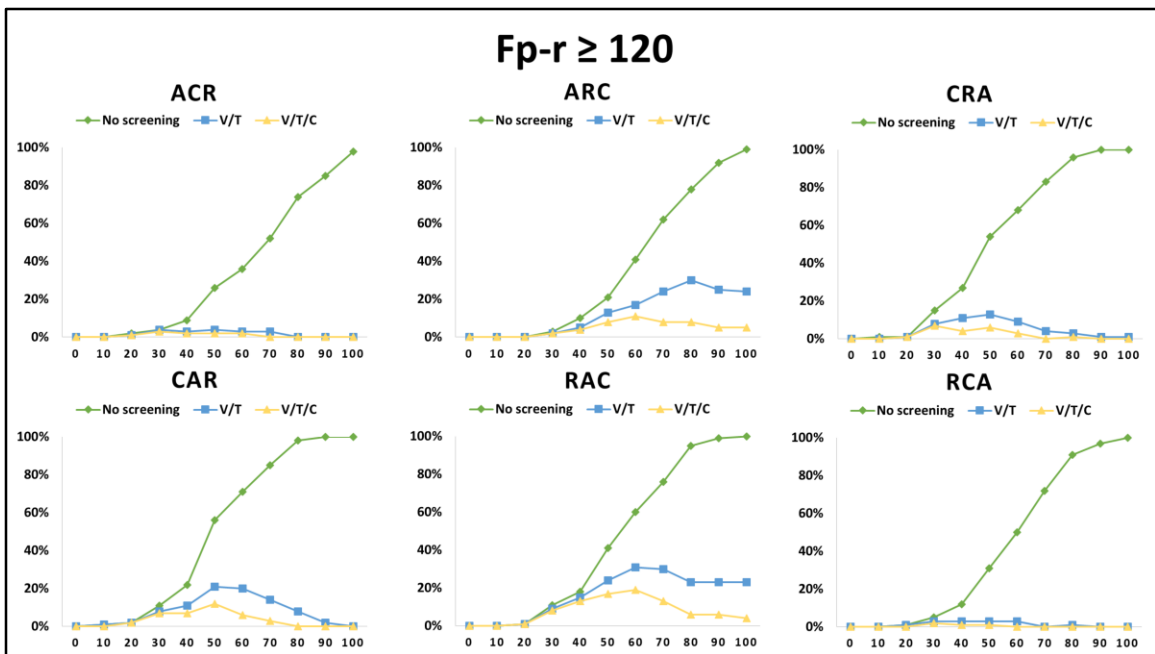
The X-axis is the percentage of mixed responding items added, and the Y-axis is the percentage of the sample elevated to the clinical cut score.

So, in this example, as we add more ARC-type mixed responding, we see a big impact on Fp (GREEN LINE). But, when we screen protocols using VRIN and TRIN (BLUE LINE), our chance of mislabeling someone as a false feigner is reduced quite a bit, and adding CRIN (YELLOW LINE) helps reduce it further.

This is a pretty extreme example, so let's take a look at the rest of the findings.

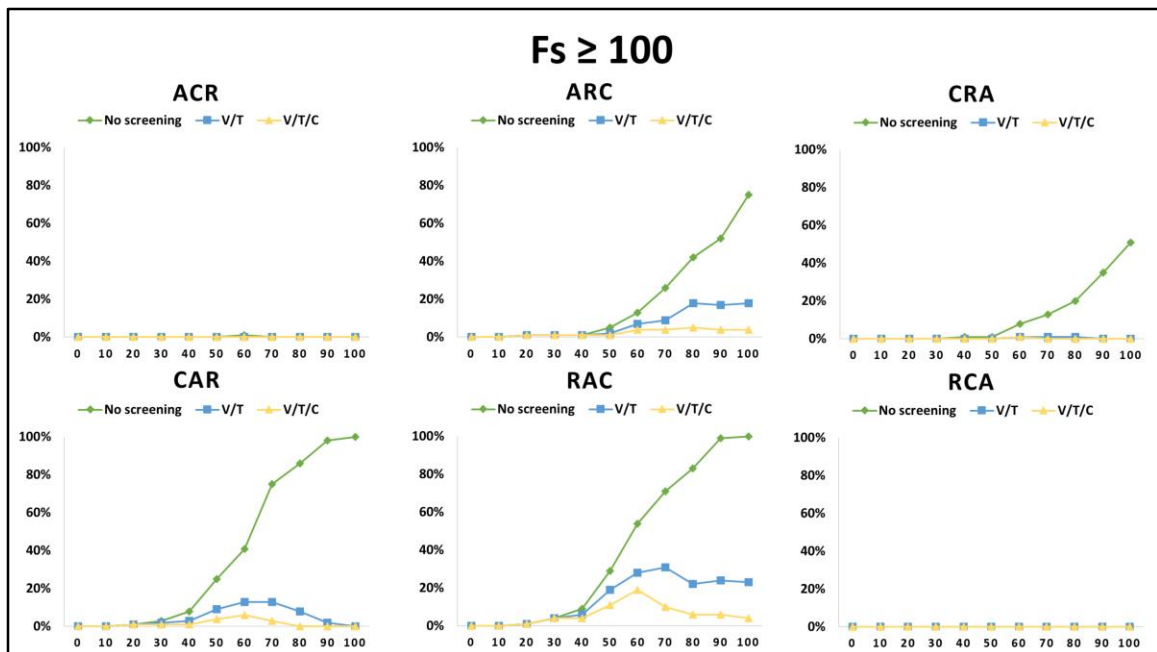


So, first, we'll take a look at how many protocols reached 120T on F-r as we add increasing levels of mixed responding. For this scale, we saw some impact in four of the six conditions. But, VRIN and TRIN flagged nearly all of those protocols, and didn't give CRIN much of a chance to help further.



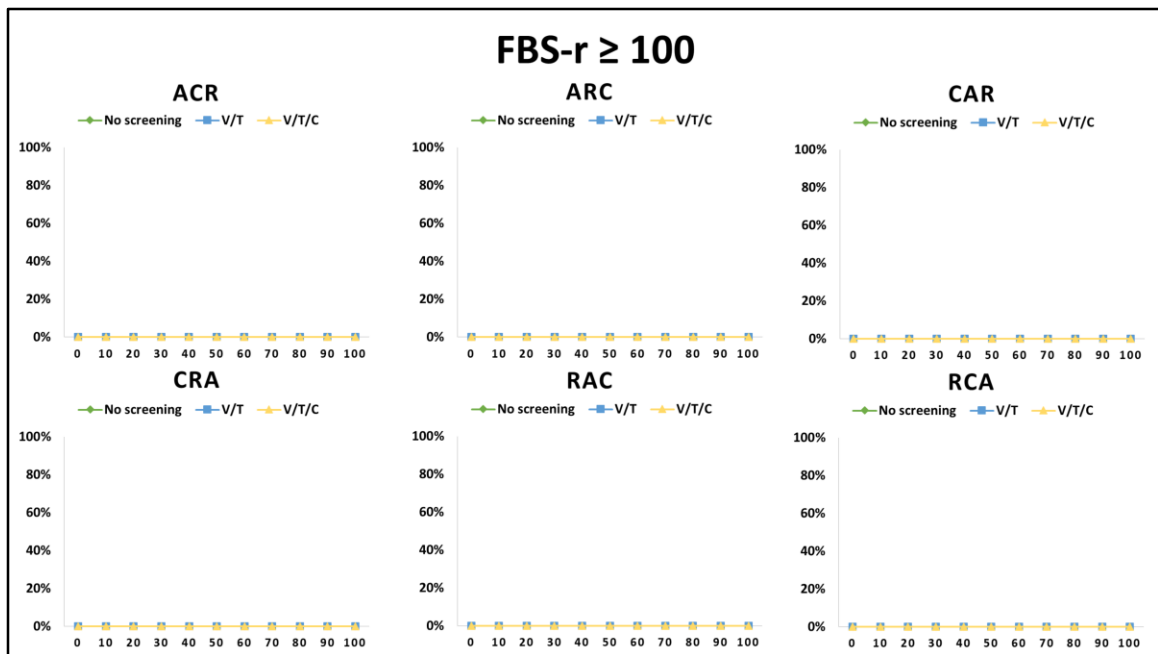
As hypothesized, Fp was very impacted by mixed responding across all six conditions. VRIN and TRIN helped out quite a bit, but in some cases – like ARC, CAR, and RAC – they left 20-30% of protocols with moderate levels of mixed responding undetected.

In these cases, CRIN was able to detect even more, dropping those rates quite a bit. However, even with CRIN, there is still some concern about false feigning rates in some conditions, like RAC.

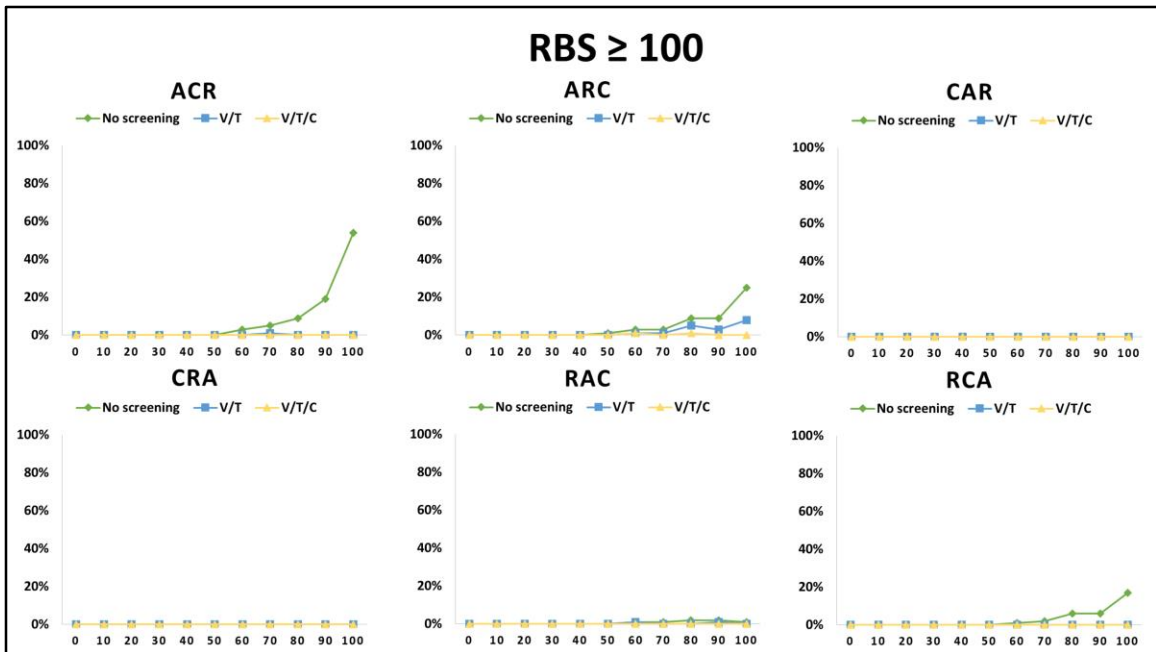


The impact on F_s depended on condition. Remember how there are several true-keyed items in the middle section? As hypothesized, we see the biggest impact where we added acquiescent (and to a lesser degree, random) data to the middle section.

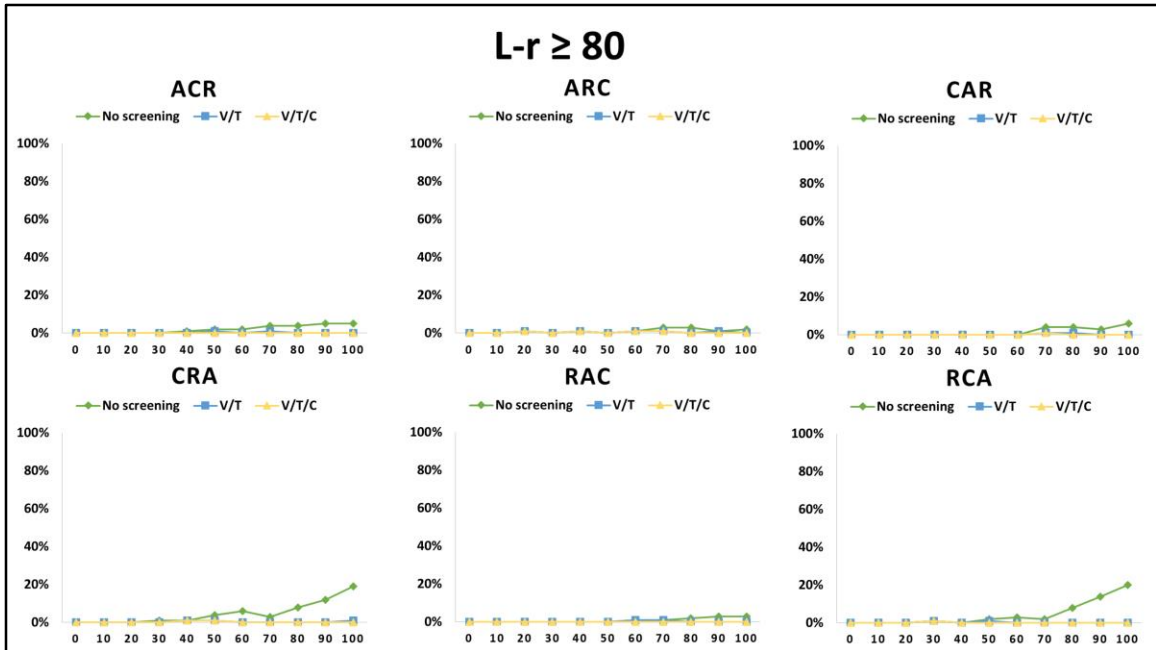
Again, VRIN and TRIN do a pretty job of flagging most of these protocols, but CRIN is incrementally useful in catching even more.



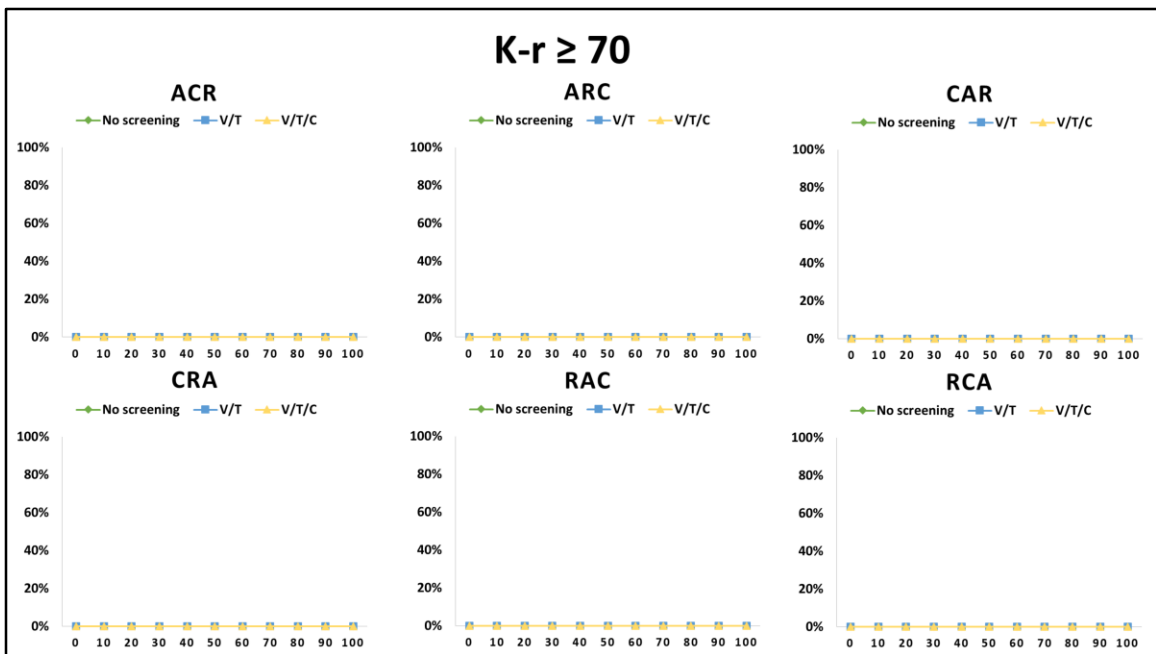
FBS, which requires 80% of items to be endorsed before it reaches 100T, wasn't so sensitive to mixed responding that it led to elevations, so not much to see here!



There was a small impact on RBS for some of the conditions, but you can see that it's well-managed by VRIN and TRIN.

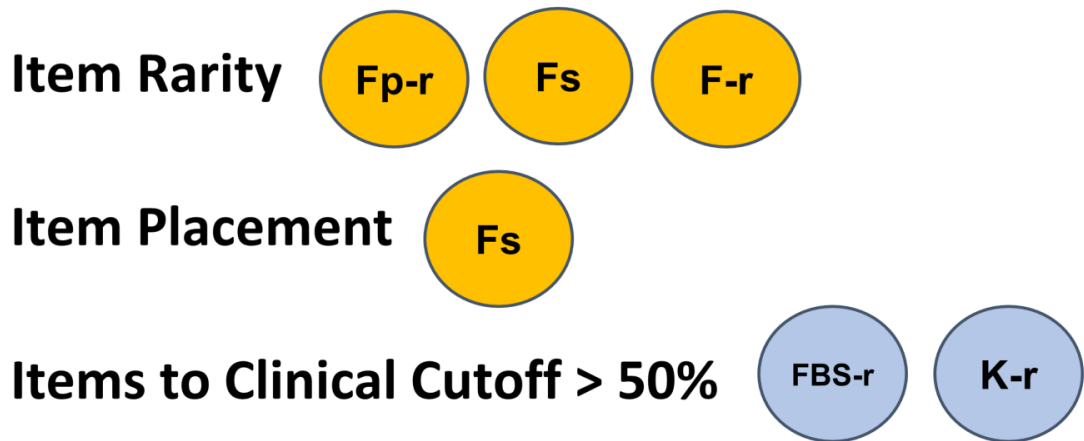


There wasn't much impact on L's elevations either, which requires 64% of items to be endorsed to reach 80T. But, VRIN and TRIN managed the few issues that popped up.



And finally, much like FBS, K (which requires 100% of items to be endorsed to reach 70T) elevations were not impacted by mixed responding at all.

Discussion



To summarize these findings:

- Scales comprised of rare items are particularly susceptible to mixed responding without the use of validity scales (i.e., Fp-r, Fs, F-r). Although these scales are intentionally designed to consist of rare items, and that makes them effective, clinicians should be aware of the potential for false positive elevations due to undetected mixed responding.
- Another factor influencing scale susceptibility is item placement as demonstrated with Fs. As much as possible, both true- and false-keyed validity scale items should be distributed throughout the test booklet.
- Those scales requiring high endorsement rates seemed most protected against mixed responding. Although difficult when developing rare-item scales, test developers should consider validity scales that require endorsement of many items to avoid the impact of mixed responding.

Discussion

TRIN-r

VRIN-r

CRIN

- Our study also showed the utility of MMPI-2-RF indicators of non-content-based invalid responding. TRIN-r and VRIN-r significantly reduce the impact of mixed responding on content-based invalid responding.
- CRIN shows incremental utility in reducing the likelihood of false feigner misinterpretations, especially for Fp and Fs.
- This is consistent with guidance that when non-content based validity scales are elevated, content-based validity scales should be interpreted with caution.
- But it extends into the realm of mixed responding and supports the usefulness of CRIN on the MMPI-2-RF and its consideration for future versions of the test.

Discussion

TRIN-r



VRIN-r



CRIN

- Our study also showed the utility of MMPI-2-RF indicators of non-content-based invalid responding. TRIN-r and VRIN-r significantly reduce the impact of mixed responding on content-based invalid responding.
- CRIN shows incremental utility in reducing the likelihood of false feigner misinterpretations, especially for Fp and Fs.
- This is consistent with guidance that when non-content based validity scales are elevated, content-based validity scales should be interpreted with caution.
- But it extends into the realm of mixed responding and supports the usefulness of CRIN on the MMPI-2-RF and its consideration for future versions of the test.

Discussion

TRIN-r



VRIN-r



CRIN



- Our study also showed the utility of MMPI-2-RF indicators of non-content-based invalid responding. TRIN-r and VRIN-r significantly reduce the impact of mixed responding on content-based invalid responding.
- CRIN shows incremental utility in reducing the likelihood of false feigner misinterpretations, especially for Fp and Fs.
- This is consistent with guidance that when non-content based validity scales are elevated, content-based validity scales should be interpreted with caution.
- But it extends into the realm of mixed responding and supports the usefulness of CRIN on the MMPI-2-RF and its consideration for future versions of the test.

Discussion

TRIN-r ✓

VRIN-r ✓

CRIN ✓

"False Feigners": Examining the Impact of Non-Content-Based Invalid Responding on the Minnesota Multiphasic Personality Inventory-2 Restructured Form Content-Based Invalid Responding Indicators

Danielle Burchett
California State University, Monterey Bay

Wendy R. Dragon
Wright State University

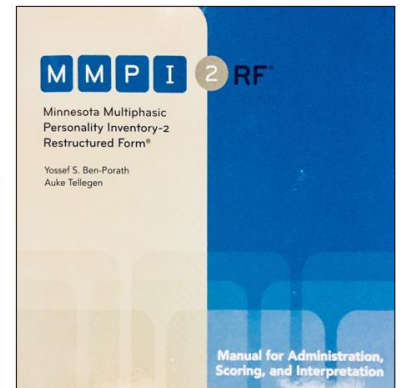
Ashley M. Smith-Holbert and
Anthony M. Tarescavage
Kent State University

Curtis A. Mattson
The Forest Institute of Professional Psychology

Richard W. Handel
Eastern Virginia Medical School

Yossef S. Ben-Porath
Kent State University

Misinterpretation of non-content based invalid (e.g., random, fixed) responding as overreporting or underreporting is likely to adversely impact test interpretation and could bias inferences about examinee intentions. We examined the impact of non-content based invalid responding on the following Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2 RF) content-based invalid responding indicators (Fp, F, Fv).



- Our study also showed the utility of MMPI-2-RF indicators of non-content-based invalid responding. TRIN-r and VRIN-r significantly reduce the impact of mixed responding on content-based invalid responding.
- CRIN shows incremental utility in reducing the likelihood of false feigner misinterpretations, especially for Fp and Fs.
- This is consistent with guidance that when non-content based validity scales are elevated, content-based validity scales should be interpreted with caution.
- But it extends into the realm of mixed responding and supports the usefulness of CRIN on the MMPI-2-RF and its consideration for future versions of the test.

Limitations	Future Directions
Computer-generated data Arbitrary definition of mixed responding Relatively small sample size for simulation design	Real-world samples Explore more mixed responding conditions Larger sample size

QUICK VERSION:

And briefly, we acknowledge computer-generated findings should be supplemented with real-world data, there are many ways you could define mixed responding, and our dataset was on the smaller side.

NOTES:

Briefly, we acknowledge the limits to generalizability any time computer-generated simulation data are used. So, future work should complement this effort with real-world samples.

We also borrowed Whitney's relatively arbitrary definition of mixed responding, and recognize there are probably many other ways this could be defined.

And finally, because of our strict exclusionary criteria, our sample was relatively small. Although Burchett et al. found very similar results for smaller and larger samples, it is ideal to replicate this work using larger data sources.

Disclosures & Acknowledgements

This research was made possible by support from a **grant** from the **University of Minnesota Press, Test Division** and support from the **California State University, Monterey Bay Undergraduate Research Opportunity Center (UROC)**.

This research was approved by the **California Human Services Agency Committee for the Protection of Human Subjects**.

The **statements and opinions expressed are those of the authors** and do not constitute the official views or the official policy of DSH-Patton, The California Department of State Hospitals, or the State of California.

We want to express our appreciation for the funding provided by the University of Minnesota Press as well as our university's Undergraduate Research Opportunities Center.

Our work was approved by the State of California's IRB.

And our statements are our own, not those of the hospital or university.

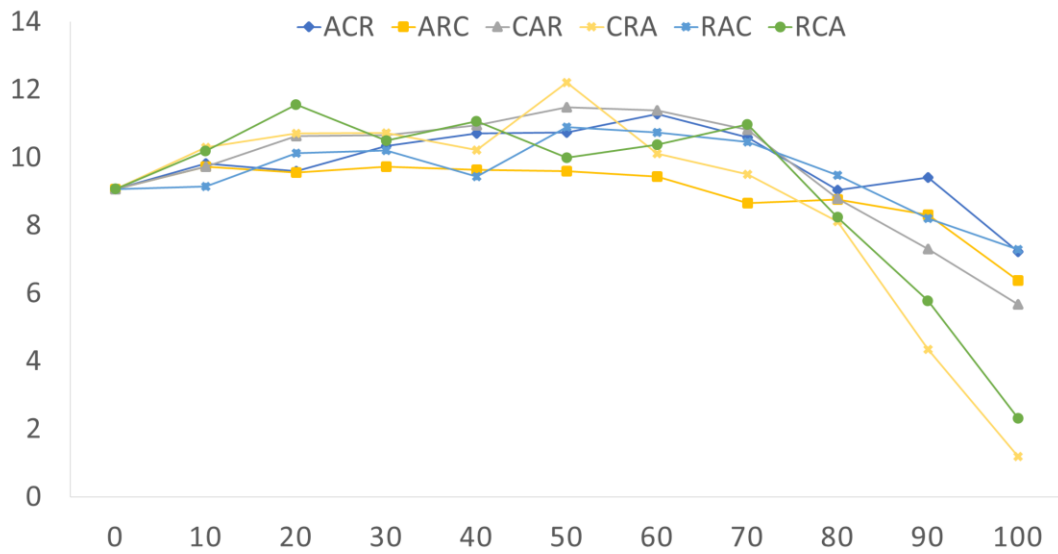
References

- Archer, R.P., Handel, R.W., Ben-Porath, Y.S., & Tellegen, A. (2016) *Minnesota Multiphasic Personality Inventory – Adolescent Restructured Form; Administration, Scoring, Interpretation and Technical Manual*. Minneapolis, MN: University of Minnesota Press.
- Burchett, D., Dragon, W. R., Smith, A. M, Tarescavage, A. M., Mattson, C. A., Ben-Porath, Y. S., & Handel, R. W. (2016). "False feigners": Examining the impact of non-content-based invalid responding on the Minnesota Multiphasic Personality Inventory-2 Restructured Form content-based invalid responding indicators. *Psychological Assessment*, 28(5), 458-470.
- Lemaster, N. M., Whitney, K., Burchett, D., & Lee, T. T. C. (2019, March). *Examining the sensitivity and incremental validity of an MMPI-2-RF Combined Response Inconsistency (CRIN) Scale for detecting mixed responding*. Paper presented at the Society for Personality Assessment, New Orleans, LA.
- Whitney, K., Chille, T., Burchett, D., Ben-Porath, Y. S., & Glassmire, D. M. (2018b, April). *Sensitivity of an MMPI-2-RF Combined Response Inconsistency (CRIN) Scale to mixed responding*. Presentation presented at the MMPI Workshops & 53rd Annual Symposium, Hollywood, FL.
- Whitney, K., Chille, T., Burchett, D., Ben-Porath, Y. S., & Glassmire, D. (2018, March). *MMPI-2-RF Combined Response Inconsistency (CRIN) basic properties in normative and forensic samples*. Poster presented at the American Psychology-Law Society, Memphis, TN.

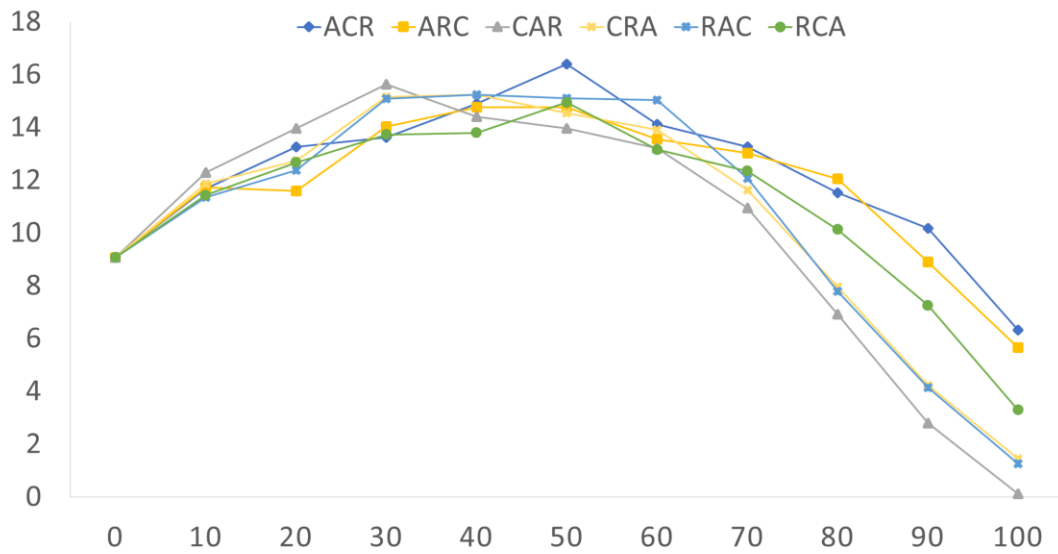
Thank You!
Questions?



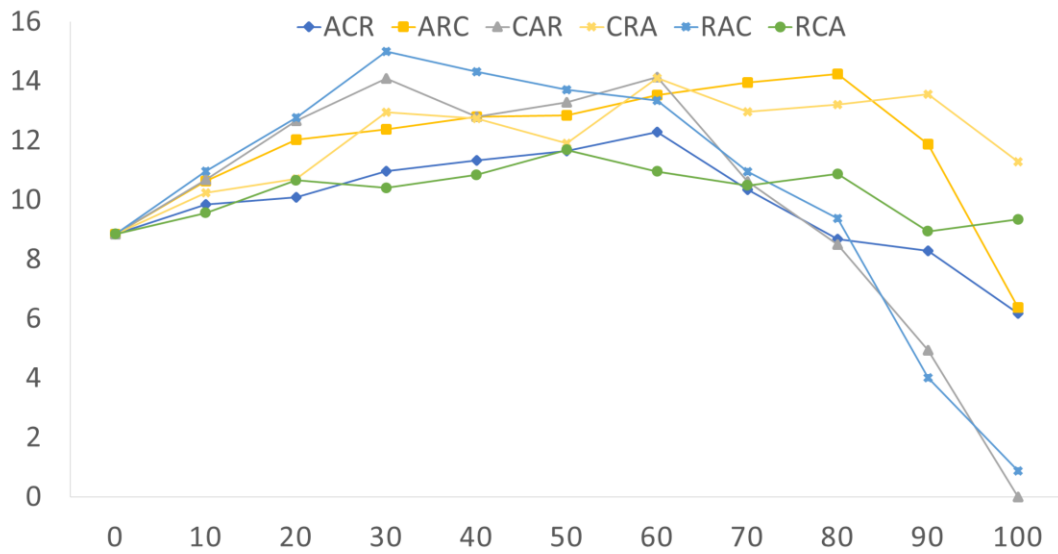
F-r Standard Deviations Increasing Mixed Responding



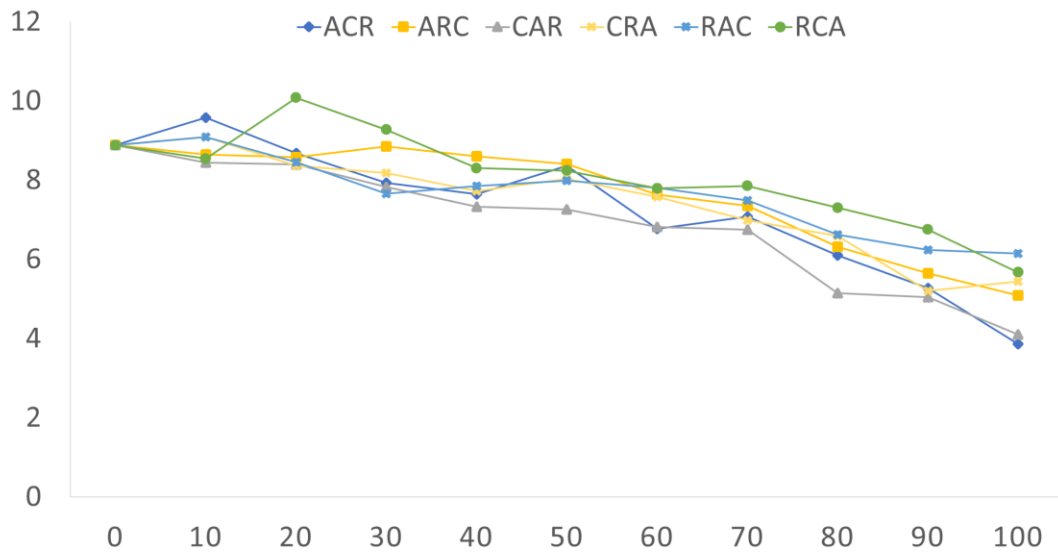
Fp-r Standard Deviations Increasing Mixed Responding



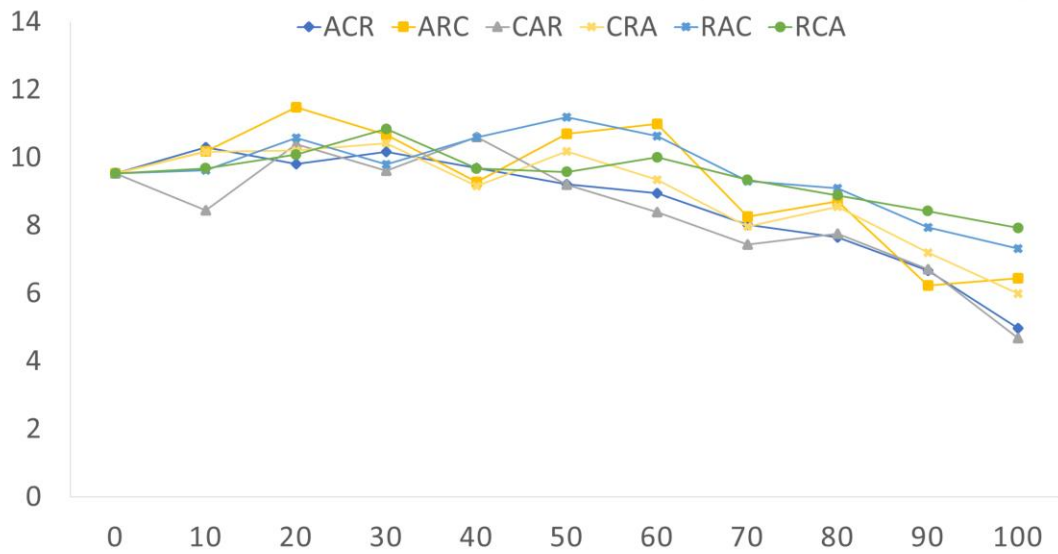
Fs Standard Deviations Increasing Mixed Responding



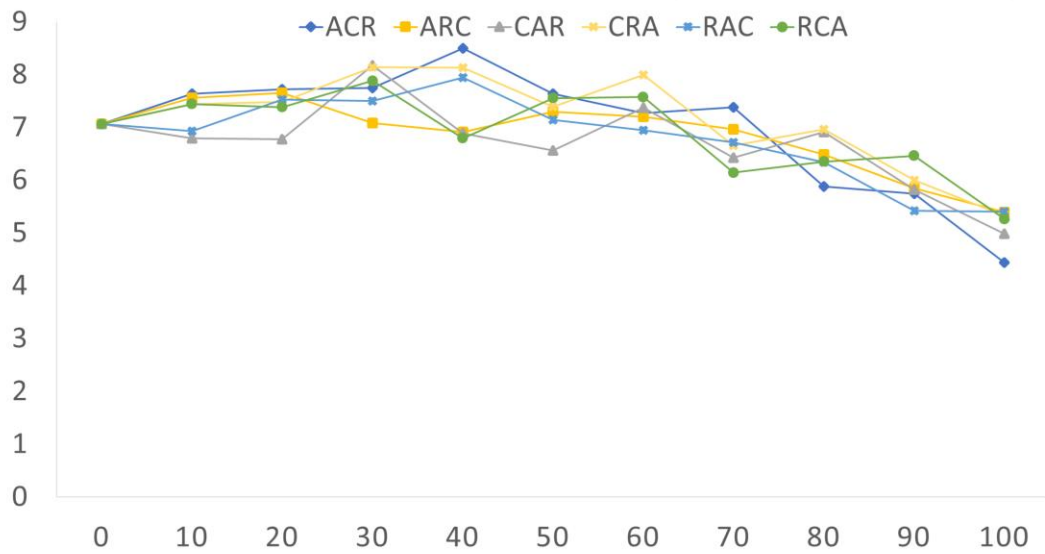
FBS-r Standard Deviations Increasing Mixed Responding



RBS Standard Deviations Increasing Mixed Responding



L-r Standard Deviations Increasing Mixed Responding



K-r Standard Deviations Increasing Mixed Responding

