Global Assessment of Functioning (GAF) Scale

Consider psychological, social, and occupational functioning on a hypothetical continuum of mental health–illness. Do not include impairment in functioning due to physical (or environmental) limitations.

Code (Note: Use intermediate codes when appropriate, e.g., 45, 68, 72.)

100 | 91 **Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities. No symptoms.**

90 | 81 **Absent or minimal symptoms** (e.g., mild anxiety before an exam), **good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, no more than everyday problems or concerns** (e.g., an occasional argument with family members).

80 | 71 **If symptoms are present, they are transient and expectable reactions to psychosocial stressors** (e.g., difficulty concentrating after family argument); **no more than slight impairment in social, occupational, or school functioning** (e.g., temporarily falling behind in schoolwork).

70 | 61 **Some mild symptoms** (e.g., depressed mood and mild insomnia) **OR some difficulty in social, occupational, or school functioning** (e.g., occasional truancy, or theft within the household), **but generally functioning pretty well, has some meaningful interpersonal relationships.**

60 | 51 **Moderate symptoms** (e.g., flat affect and circumstantial speech, occasional panic attacks) **OR moderate difficulty in social, occupational, or school functioning** (e.g., few friends, confli , with peers or co-workers).

50 | 41 **Serious symptoms** (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) **OR any serious impairment in social, occupational, or school functioning** (e.g., no friends, unable to keep a job).

40 | 31 **Some impairment in reality testing or communication** (e.g., speech is at times illogical, obscure, or irrelevant) **OR major impairment in several areas, such as work or school, family relations, judgment, thinking, or mood** (e.g., depressed man avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school).

30 | 21 **Behavior is considerably influenced by delusions or hallucinations OR serious impairment in communication or judgment** (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) **OR inability to function in almost all areas** (e.g., stays in bed all day; no job, home, or friends).

20 | 11 **Some danger of hurting self or others** (e.g., suicide attempts without clear expectation of death; frequently violent; manic excitement) **OR occasionally fails to maintain minimal personal hygiene** (e.g., smears feces) **OR gross impairment in communication** (e.g., largely incoherent or mute).

10 | 1 **Persistent danger of severely hurting self or others** (e.g., recurrent violence) **OR persistent inability to maintain minimal personal hygiene OR serious suicidal act with clear expectation of death.**

0 Inadequate information.

The rating of overall psychological functioning on a scale of 0–100 was operationalized by Luborsky in the Health-Sickness Rating Scale (Luborsky L: "Clinicians' Judgments of Mental Health." *Archives of General Psychiatry* 7:407–417, 1962). Spitzer and colleagues developed a revision of the Health-Sickness Rating Scale called the Global Assessment Scale (GAS) (Endicott J, Spitzer RL, Fleiss JL, Cohen J: "The Global Assessment Scale: A Procedure for Measuring Overall Severity of Psychiatric Disturbance." *Archives of General Psychiatry* 33:766–771, 1976). A modified version of the GAS was included in DSM-III-R as the Global Assessment of Functioning (GAF) Scale.

TABLE 1. DIAGNOSTIC TEST CHARACTERISTICS OF THE MINI-MENTAL STATE EXAMINATION AMONG THREE EDUCATIONAL STRATA

	Educational Attainment		
	Middle School	High School	College/ Graduate School
Number Demented:	23/40 (58%)	33/63 (52%)	53/107 (50%)
ROC Curve Area:	.95	.95	.96
MMSE Threshold*	Sens/Spec**	Sens/Spec**	Sens/Spec**
19	.61/.94	.51/1.00	.45/1.00
20	.65/.94	.58/.97	.55/1.00
21	.82/.94	.58/.97	.68/1.00
22	.82/.88	.70/.97	.75/1.00
23	1.00/.71	.79/.97	.79/1.00
24	1.00/.59	.88/.79	.83/1.00
25	1.00/.35	1.00/.69	.87/.80
26	1.00/.24	1.00/.59	.94/.70
27	1.00/.18	1.00/.41	.98/.60
28	1.00/.06	1.00/.28	1.00/.20

* *Minimum normal score.*
** *Sensitivity/Specificity.*

TABLE 2. ACCURACY OF THE MINI-MENTAL STATE EXAMINATION FOR DETECTING DEMENTIA*

MMSE Threshold**	Educational Attainment		
	Middle School	High School	College/ Graduate School
19	.88	.90	.89
20	.88	.89	.91
21	.92	.89	.94
22	.87	.92	.95
23	.76	.93	.96
24	.67	.81	.97
25	.48	.75	.81
26	.39	.67	.75
27	.34	.53	.67
28	.25	.42	.40

* *Proportion of patients correctly classified as demented or not demented. Data assume the prevalence of dementia is 20%.*
** *Minimum normal MMSE score.*

TABLE 3. CORRELATION OF YEARS OF EDUCATION WITH MMSE SUBSECTION SCORES

	r*
Orientation (Time)	.14**
Orientation (Place)	.11
Registration	.09
Attention and calculation	.21***
Recall	.23#
Language	.19***
Total score	.23#

* *Pearson correlation coefficient. Positive values indicate that higher education is associated with higher MMSE scores, and vice versa for lower education.*
** $P < 0.05$.
*** $P < 0.01$.
\# $P < 0.001$.

DISCUSSION

An increasing and, we believe, justifiable emphasis has been placed in recent years on the use of standardized screening instruments for the detection of cognitive dysfunction and dementia in the elderly.[25–27] Given the prevalence and significance of dementia, however, the diagnostic accuracy of such instruments has considerable personal and public health ramifications. False positive results are likely to precipitate unnecessary emotional distress in patients and families as well as expensive and potentially complicated diagnostic testing and treatment. False negative results may be as consequential if reversible or remediable causes of dementia are not recognized and treated. Thus, such instruments should be carefully calibrated to the populations in which they are used.

Previous studies have noted associations between education and MMSE scores and questioned the validity of the MMSE in poorly educated persons.[9–12] We used decision analytic techniques to optimize MMSE norms and evaluate its accuracy in various educational groups. These results indicate the MMSE is an accurate screening test for Alzheimer's dementia among both less well and highly educated older adults if education-specific norms are applied. These results also suggest the lack of MMSE specificity noted previously in poorly educated persons appears not to reflect an inherent lack of accuracy in the MMSE in such populations. Rather, it appears to be an artifact related to subjecting poorly educated individuals to conventional MMSE norms. When lower norms are applied, the MMSE appears to be highly accurate in persons with middle school education. However, the accuracy of lower norms in more poorly educated persons will need to be determined in subsequent studies.

MINI-MENTAL STATE (FOLSTEIN) Examiner: _____

NAME_____ SS#_____ DATE_____

Check box if correct response given. Record the incorrect response.

I. ORIENTATION: Ask the following questions. (Maximum score is 10)

			Incorrect Response:
What is today's date?	Date (e.g. Jan. 21)	☐	_____
What is the year?	Year	☐	_____
What is the month?	Month	☐	_____
What day is today?	Day (e.g. Monday)	☐	_____
Can you also tell me what season it is?	Season	☐	_____
Can you also tell me the name of this hospital (clinic)?	Hospital (Clinic)	☐	_____
What floor are we on?	Floor	☐	_____
What town or city are we in?	Town or City	☐	_____
What county are we in?	County	☐	_____
What state are we in?	State	☐	_____

Subscore_____

II. IMMEDIATE RECALL: Circle list used. (Maximum score is 3)

Baby	Daughter	Village	Ball	Apple	☐	_____
Garden	River	Heaven	Flag	Penny	☐	_____
Leader	Table	Finger	Tree	Table	☐	_____

Subscore_____

III. ATTENTION AND CALCULATION: (Maximum score is 5)

93		D		☐ _____
86		L		☐ _____
79	or	R	(If patient refuses	☐ _____
72		O	to subtract)	☐ _____
65		W		☐ _____

Subscore_____

IV. RECALL: (of above list used) (Maximum score is 3)

☐ _____
☐ _____
☐ _____

Subscore_____

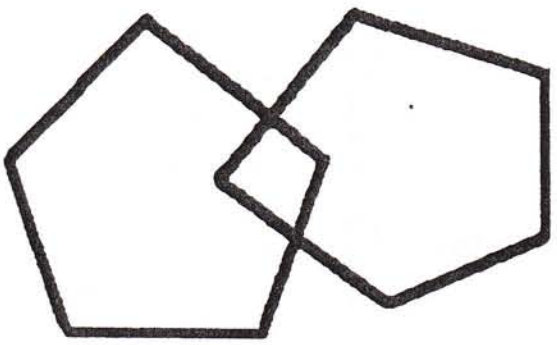
V. LANGUAGE: (Maximum score is 9)

NAMING:	Watch	☐ _____
	Pen	☐ _____
REPETITION: "No ifs, ands or buts".		☐ _____
3-STAGE COMMAND: Give the subject a piece of plain blank paper and say, "Take the paper in your right hand, fold it in half with both hands and place it in your lap".	Takes with right hand	☐ _____
	Folds paper in half	☐ _____
	Puts paper on lap	☐ _____
READING: Score correctly only if he/she actually closes eyes.		☐ _____
WRITING: Have the subject write a complete sentence		☐ _____
COPYING: Ask the subject to copy the intersecting pentagons		☐ _____

Subscore_____

TOTAL SCORE: (Maximum score is 30.) TOTAL_____

CLOSE YOUR EYES



Mini-Mental Status Examination (MMSE), Montreal Cognitive Assessment (MoCA), and the Saint Louis Mental Status Examination (SLUMS)

The MMSE was the widely used default test for years, but has been removed from the public domain. Can you still use it legally? Here is from the PAR website:

> **Q: Does the administration of the MMSE in a clinical setting constitute copyright infringement?**
> **A:** No. As long as the MMSE is not copied or reproduced, the administration of the test does not constitute copyright infringement. Hence, if a person has an authorized (legal) version of the MMSE (a copy that was not illegally obtained or produced) or has it memorized and administers the test, there has been no copyright infringement. Answers and scores may be recorded. Please note two important caveats: 1. we should not copy (infringe on the copyright of) the official answer sheet being distributed by PAR; 2. Administering any standardized assessment instrument from memory may impact the quality of the administration, and therefore the results. Thus, caution should be taken before embarking upon administration strictly from memory.

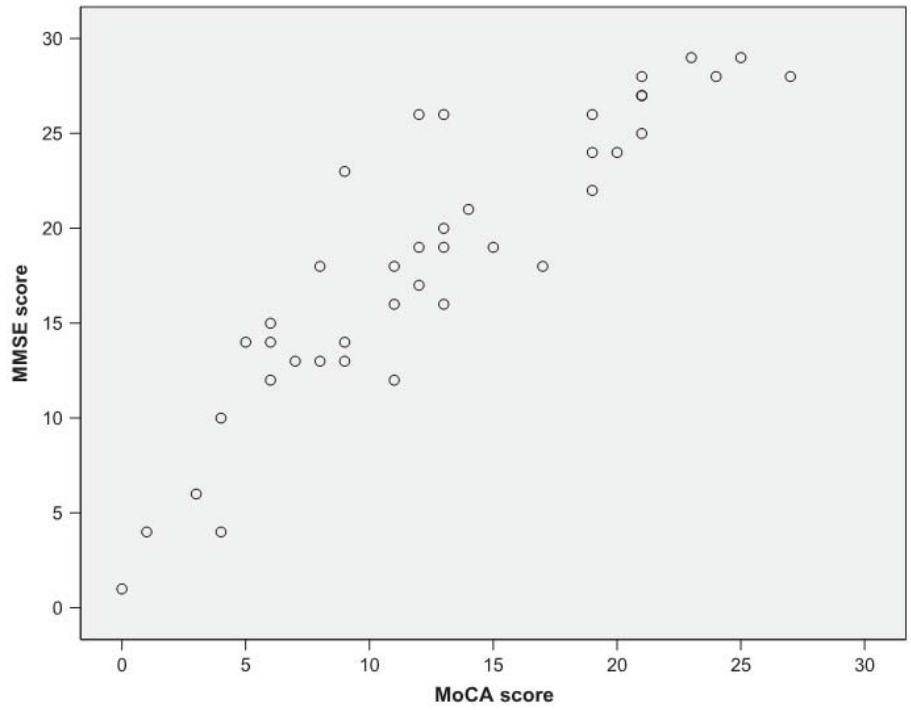


FIGURE 1 Scatterplot of MoCA against MMSE scores.

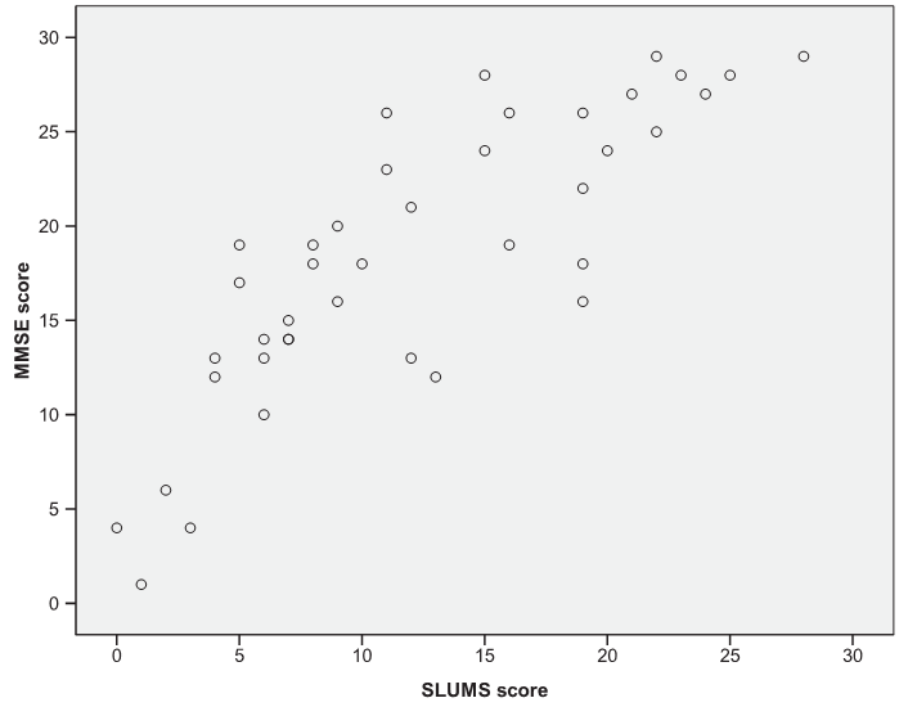


FIGURE 2 Scatterplot of SLUMS against MMSE scores.

From Stewart et al. (2012), Clinical Gerontologist, 35:57–75

VAMC
SLUMS Examination

Questions about this assessment tool? E-mail aging@slu.edu.

Name _____ Age _____

Is patient alert? _____ Level of education _____

__/1 (1) **1. What day of the week is it?**

__/1 (1) **2. What is the year?**

__/1 (1) **3. What state are we in?**

4. Please remember these five objects. I will ask you what they are later.

Apple Pen Tie House Car

5. You have $100 and you go to the store and buy a dozen apples for $3 and a tricycle for $20.

(1) **How much did you spend?**

__/3 (2) **How much do you have left?**

6. Please name as many animals as you can in one minute.

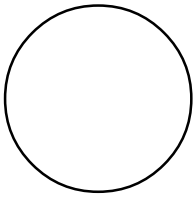
__/3 (0) 0-4 animals (1) 5-9 animals (2) 10-14 animals (3) 15+ animals

__/5 **7. What were the five objects I asked you to remember?** 1 point for each one correct.

8. I am going to give you a series of numbers and I would like you to give them to me backwards. For example, if I say 42, you would say 24.

__/2 (0) 87 (1) 649 (1) 8537

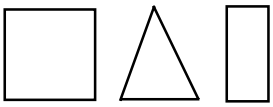
9. This is a clock face. Please put in the hour markers and the time at ten minutes to eleven o'clock.



(2) Hour markers okay

__/4 (2) Time correct

__/2 (1) **10. Please place an X in the triangle.**



(1) **Which of the above figures is largest?**

11. I am going to tell you a story. Please listen carefully because afterwards, I'm going to ask you some questions about it.

Jill was a very successful stockbroker. She made a lot of money on the stock market. She then met Jack, a devastatingly handsome man. She married him and had three children. They lived in Chicago. She then stopped work and stayed at home to bring up her children. When they were teenagers, she went back to work. She and Jack lived happily ever after.

(2) **What was the female's name?** (2) **What work did she do?**

__/8 (2) **When did she go back to work?** (2) **What state did she live in?**

_____ **TOTAL SCORE**

Department of Veterans Affairs

SAINT LOUIS UNIVERSITY

AGING SUCCESSFULLY

Scoring

High School Education		Less than High School Education
27-30	Normal	25-30
21-26	MNCD*	20-24
1-20	Dementia	1-19

* Mild Neurocognitive Disorder

SH Tariq, N Tumosa, JT Chibnall, HM Perry III, and JE Morley. The Saint Louis University Mental Status (SLUMS) Examination for Detecting Mild Cognitive Impairment and Dementia is more sensitive than the Mini-Mental Status Examination (MMSE) - A pilot study. J am Geriatri Psych (in press).

general assessment series 8

try this:®

Best Practices in Nursing Care to Older Adults

From The Hartford Institute for Geriatric Nursing, New York University, College of Nursing

Issue Number 3.2, Revised 2012

Series Editor: Marie Boltz, PhD, GNP-BC
Series Co-Editor: Sherry A. Greenberg, MSN, GNP-BC
New York University College of Nursing

Mental Status Assessment in Older Adults: Montreal Cognitive Assessment: MoCA Version 7.1 (Original Version)

By: Deirdre M. Carolan Doerflinger, CRNP, PhD
Inova Fairfax Hospital, Falls Church, VA

WHY: The incidence of mild cognitive impairment (MCI) increases with age ranging from 7% to 38% (2011 Alzheimer's disease Facts and Figures). Older adults with MCI have as high as 14% higher risk of developing Alzheimer's dementia (2011 Alzheimer's disease Facts and Figures). While studies have shown that treatment with an acetylcholinesterase inhibitor prior to progression has delayed dementia onset by 3 years, currently there is no endorsed treatment recommendations for MCI.

BEST TOOL: The Montreal Cognitive Assessment (MoCA© Version 7.1) was developed as a quick screening tool for MCI and early Alzheimer's dementia. It assesses the domains of attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. There are two alternative MoCA© forms (Version 7.2 and 7.3) available in an effort to decrease possible learning effects when used repeatedly (Phillips et al., 2011). The MoCA© has been tested extensively for use in a variety of disorders affecting cognition such as HIV, Huntington's chorea, Multiple Sclerosis, Parkinson's disease, stroke, vascular dementia, and substance abuse in addition to the well older adult. It has been tested in 14 different languages, ages ranging from as young as 49 in two reports to old-old (85+) with a variety of education levels. The total possible score is 30 points with a score of 26 or more considered normal. To better adjust the MoCA for lower educated individuals, 2 points should be added to the total MoCA score for those with 4-9 years of education and 1 point for 10-12 years of education (Johns et al., 2010). The score range for MCI is 19-25.2 and for Alzheimer's dementia 11.4-21. While the score ranges overlap, differentiation between the conditions is dependent upon associated functional impairment. A modified version, MoCA-B, has been developed for use in visual impairments.

TARGET POPULATION: The MoCA can be used in a variety of settings from primary care to acute care. It may be used in culturally diverse populations, a variety of ages and differing educational levels.

VALIDITY AND RELIABILITY: The MoCA detected MCI with 90%-96% range sensitivity and specificity of 87% with 95% confidence interval. The MoCA detected 100% of Alzheimer's dementia with a specificity of 87%.

STRENGTHS AND LIMITATIONS: The MoCA takes approximately 10 minutes to administer. It is accessible via the MoCA© website, http://www.mocatest.org/ with clear administration and scoring instructions (refer to website for copyright information). All these items, test, instructions and scoring are available in 36 languages. There is some recent research suggesting that lowering the threshold score to 23 may prevent over identification of normal individuals. It has been tested in a variety of settings and populations and displayed accuracy in identification of MCI and Alzhiemer's dementia.

FOLLOW-UP: The U.S. Preventative Services Task Force in 2003, made no formal recommendations for screening for dementia. The American Academy of Neurology (2001) determined that there is not sufficient evidence to recommend cognitive screening of asymptomatic individuals. This guideline is currently under revision. The American Medical Association (2003) and the American Academy of Family Physicians (2001) recommend that health care providers be alert for cognitive and functional decline in elderly patients for recognition of dementia in its early stages. Annual screening, as a component of the annual physical, is realistic.

MORE ON THE TOPIC:

Best practice information on care of older adults: www.ConsultGeriRN.org.
MoCA website: http://www.mocatest.org/.

2011 Alzheimer's Facts and Figures. Washington DC: Alzheimer's Association. No. 7. Accessed September 18, 2011 from http://www.alz.org/downloads/Facts_Figures_2011.pdf.
Berstein, I.H., Lacritz, L., Barlow, C.F., Weiner, M.F., & DeFina, L.F. (2011). Psychometric evaluation of the Montreal Cognitive Assessment (MoCA) in three diverse samples. *The Clinical Neuropsychologist*, *25*(1), 119-126.
Dalrymple-Alford, J., MacAskill, M., Nakas, C., et al. (2010). The MoCA: Well-suited screen for cognitive impairment in Parkinson 's disease. *Neurology*, *75*, 1717.1725.
Dong, Y., Sharma, V., Chan, B., et al. (2010). The Montreal Cognitive Assessment (MoCA) is superior to the Mini-Mental State Examination (MMSE) for the detection of vascular cognitive impairment after acute stroke. *Journal of Neurological Sciences*, *299*, 15-18.
Johns, E.K. et al. Level of education and performance on the Montreal Cognitive Assessment (MoCA©): New recommendations for education corrections. Presented at the Cognitive Aging Conference 2010, Atlanta, Georgia, April 15-18th, 2010.
McLennan, S., Mathias, J., Brennan, L., & Stewart, S. (2011). Validity of the Montreal Cognitive Assessment (MoCA) as a screening test for mild cognitive impairment (MCI) in a cardiovascular population. *Journal of Geriatrics Psychiatry*, *24*, 33-38.
Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *JAGS*, *53*, 695–699.
Phillips, N. et al. Validation of alternate forms for the Montreal Cognitive Assessment (MoCA©). Presented at the 39th International Neuropsychological Society Meeting in Boston February 2-5, 2011.
Wittich, W., Phillips, N., Nasreddine, Z., & Chertkow, H. (2010). Sensitivity and specificity of the Montreal Cognitive Assessment modified for individuals who are visually impaired. *Journal of Visual Impairment & Blindness, 104*(6), 360-368.

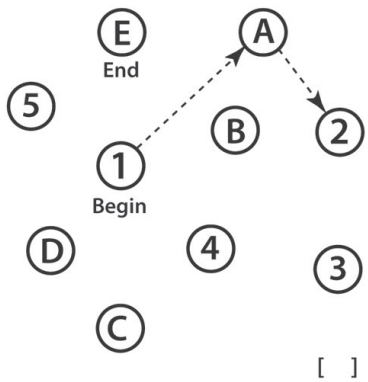
Permission is hereby granted to reproduce, post, download, and/or distribute, this material *in its entirety only* for not-for-profit educational purposes only, provided that The Hartford Institute for Geriatric Nursing, New York University, College of Nursing is cited as the source. This material may be downloaded and/or distributed in electronic format, including PDA format. Available on the internet at **www.hartfordign.org** and/or **www.ConsultGeriRN.org**. E-mail notification of usage to: **hartford.ign@nyu.edu**.

MONTREAL COGNITIVE ASSESSMENT (MOCA)

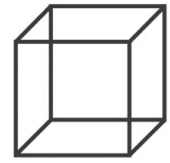
Version 7.1 Original Version

NAME :
Education : Date of birth :
Sex : DATE :

VISUOSPATIAL / EXECUTIVE



[ ]



Copy cube

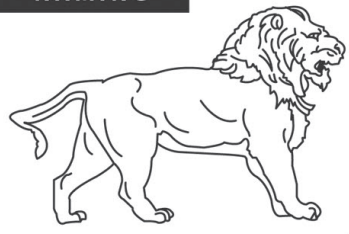
[ ]

Draw CLOCK (Ten past eleven)
(3 points)

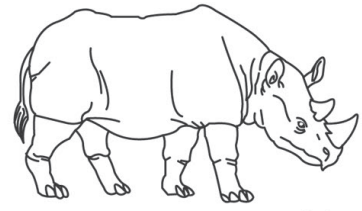
[ ] Contour [ ] Numbers [ ] Hands

POINTS ___/5

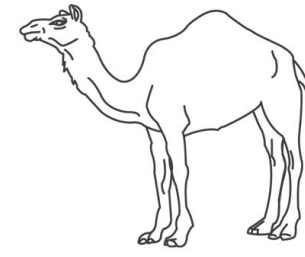
NAMING



[ ]



[ ]



[ ]

___/3

MEMORY

Read list of words, subject must repeat them. Do 2 trials, even if 1st trial is successful. Do a recall after 5 minutes.

	FACE	VELVET	CHURCH	DAISY	RED
1st trial					
2nd trial					

No points

ATTENTION

Read list of digits (1 digit/ sec.).

Subject has to repeat them in the forward order [ ] 2 1 8 5 4

Subject has to repeat them in the backward order [ ] 7 4 2

___/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors

[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B

___/1

Serial 7 subtraction starting at 100 [ ] 93 [ ] 86 [ ] 79 [ ] 72 [ ] 65

4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt**

___/3

LANGUAGE

Repeat : I only know that John is the one to help today. [ ]

The cat always hid under the couch when dogs were in the room. [ ]

___/2

Fluency / Name maximum number of words in one minute that begin with the letter F [ ] _____ (N ≥ 11 words)

___/1

ABSTRACTION

Similarity between e.g. banana - orange = fruit [ ] train – bicycle [ ] watch - ruler

___/2

DELAYED RECALL

	Has to recall words WITH NO CUE	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]	Points for UNCUED recall only
Optional	Category cue						
	Multiple choice cue						

___/5

ORIENTATION

[ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City

___/6

© Z.Nasreddine MD

www.mocatest.org

Normal ≥ 26 / 30

TOTAL ___/30

Add 1 point if ≤ 12 yr edu

Administered by: _____

Copyright© Dr Ziad S. Nasreddine, MD, FRCP — The Montreal Cognitive Assessment — MoCA© — McGill University, and Sherbrooke University Canada. Reproduced with permission. Copies are available at www.mocatest.org.

try this: general assessment series
Best Practices in Nursing Care to Older Adults

A series provided by The Hartford Institute for Geriatric Nursing, New York University, College of Nursing

EMAIL hartford.ign@nyu.edu HARTFORD INSTITUTE WEBSITE www.hartfordign.org
CLINICAL NURSING WEBSITE www.ConsultGeriRN.org

U.S.A. SUICIDE: 2019 OFFICIAL FINAL DATA

	Number	Per Day	Rate	% of Deaths
Nation	47,511	130.2	14.5	1.7
Males	37,256	102.1	23.0	2.5
Females	10,255	28.1	6.2	0.7
Whites	41,935	114.9	16.4	1.7
Nonwhites	5,576	15.3	7.6	1.2
Blacks/African American	3,309	9.1	7.1	0.9
Older Adults (65+ yrs.)	9,173	25.1	17.0	0.4
Young (15-24 yrs.)	5,954	16.3	13.9	20.0
Middle Aged (45-64 yrs.)	16,250	44.5	19.5	3.0

Group (Number of Suicides)	Rate
White Male (32,964)	26.1
White Female (8,971)	7.0
Nonwhite Male (4,292)	12.2
Nonwhite Female (1,284)	3.4
Black/African American Male (2,638)	11.8
Black/African American Female (671)	2.8
Hispanic/Latino (4,331)	7.2
Native Americans/Alaska Natives (658)	13.8
Asian/Pacific Islanders (1,609)	7.4

***Fatal Outcomes (Suicides)*:** *a 2% rate decrease was seen from 2018 to 2019, the first decrease observed in the US since a 0.36% rate decrease from 2004 to 2005*

- Average of 1 person every 11.1 minutes killed themselves—1 male every 14.1 minutes, 1 female every 51.3 minutes
- Average of 1 older adult every 57.3 minutes killed themselves; Average of 1 middle aged adult every 32.3 minutes
- Average of 1 young person every 1 hour and 28.3 minutes killed themselves. (If the 546 suicides below age 15 are included, 1 young person every 1 hour and 20.9 minutes)
- 10th ranking cause of death in U.S.— 2nd for **young**
- 3.6 male deaths by suicide for each female death by suicide
- Suicide ranks 10th as a cause of death; Homicide ranks 16th

Leading Causes of Death 15-24 yrs

Cause	Number	Rate
All Causes	29,771	69.7
1-Accidents	11,755	27.5
2-Suicide	5,954	13.9
3-Homicide	4,774	11.2
10-14 yrs	534	2.6
15-19 yrs	2,210	10.5
20-24 yrs	3,744	17.3

Nonfatal Outcomes (Attempt Survivors§) (figures are estimates)*:*

- 1,187,775 annual attempts in U.S. (using 25:1 ratio); 2019 SAMHSA study: 1.4 million adults (18 and up)
- Translates to one attempt every 26.6 seconds (based on 1,187,775 attempts) [1.4 million = 1 every 23 seconds]
- 25 attempts for every death by suicide for nation (one estimate); 100-200:1 for young; 4:1 for older adults
- 3 female attempts for each male attempt

Postvention (Exposure and Survivors of Suicide Loss)

Exposed ("Affected") – those who "know" someone personally who has died by suicide † (figures are estimates)

°Recent (Cerel, 2015) research-based estimate suggests that for each death by suicide →147 people are *exposed* (for 2019, 6.98 million annually) – among the exposed there are subgroups with a variety of effect levels (see Cerel et al., 2014) – as many as 40-50% of the population have been exposed to suicide in their lifetime based on a 2016 representative sample's results (Feigelman et al., 2017)

Suicide Loss Survivors (those bereaved of suicide - definition below): † (figures are estimates) [Subgroup of "Exposed" above]

°Survivors of Suicide Loss = experience high levels of distress for a considerable length of time after exposure (Jordan & McIntosh, 2011)

°Among those exposed to a death by suicide, *more than* 6 experience a major life disruption (loss survivors; a low, non-research based estimate see Cerel et al. 2015)

- *If each suicide has devastating effects and intimately affects > 6 other people, there are* ***over 285,000 loss survivors a year***
- Based on the 916,115 suicides from 1995 through 2019, therefore, the number of *survivors of suicide loss* in the U.S. is *more than* 5.4 million (1 of every 60 Americans in 2019); number grew by more than 285,066 in 2019
- If there is a suicide every 11.1 minutes, then there are *more than* 6 new loss survivors every 11.1 minutes as well

Suicide Methods	Number	Rate	Percent of Total		Number	Rate	Percent of Total
Firearm suicides (1st)	23,941	7.3	50.4%	All but Firearms	23,570	7.2	49.6%
Suffocation/Hanging (2nd)	13,563	4.1	28.5%	Poisoning (3rd)	6,125	1.9	12.9%
Cut/pierce (5th)	921	0.3	1.9%	Drowning (7th)	506	0.2	1.1%

U.S.A. Suicide Rates 2009-2019

(Rates per 100,000 population)

Group/ Age	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Group/ Age
5-14	0.6	0.7	0.7	0.8	1.0	1.0	1.0	1.1	1.3	1.5	1.3	5-14
15-24	10.0	10.5	11.0	11.1	11.1	11.6	12.5	13.2	14.5	14.5	13.9	15-24
25-34	13.1	14.0	14.6	14.7	14.8	15.1	15.7	16.5	17.5	17.6	17.5	25-34
35-44	16.1	16.0	16.2	16.7	16.2	16.6	17.1	17.4	17.9	18.2	18.1	35-44
45-54	19.2	19.6	19.8	20.0	19.7	20.2	20.3	19.7	20.2	20.0	19.6	45-54
55-64	16.4	17.5	17.1	18.0	18.1	18.8	18.9	18.7	19.0	20.2	19.4	55-64
65-74	13.7	13.7	14.1	14.0	15.0	15.6	15.2	15.4	15.6	16.3	15.5	65-74
75-84	15.8	15.7	16.5	16.8	17.1	17.5	17.9	18.2	18.0	18.7	18.6	75-84
85+	16.4	17.6	16.9	17.8	18.6	19.3	19.4	19.0	20.1	19.1	20.1	85+
65+	14.8	14.9	15.3	15.4	16.1	16.7	16.6	16.7	16.8	17.4	17.0	65+
Total	12.0	12.4	12.7	12.9	13.0	13.4	13.7	13.9	14.5	14.8	14.5	Total
Men	19.3	19.9	20.2	20.6	20.6	21.1	21.5	21.8	22.9	23.4	23.0	Men
Women	5.0	5.2	5.4	5.5	5.7	6.0	6.2	6.2	6.3	6.4	6.2	Women
White	13.7	14.1	14.5	14.7	14.9	15.5	15.8	15.9	16.5	16.8	16.4	White
Nonwh	5.6	5.8	5.8	6.1	6.0	6.0	6.3	6.8	7.2	7.5	7.6	NonWh
Black	5.0	5.1	5.3	5.5	5.4	5.5	5.6	6.1	6.7	7.0	7.1	Black
45-64	17.9	18.6	18.6	19.1	19.0	19.5	19.6	19.2	19.6	20.1	19.5	45-64

15 Leading Causes of Death in the U.S.A., 2019

(total of 2,854,838 deaths; 869.7 rate)

Rank & Cause of Death	Rate	Deaths
1 Diseases of heart (heart disease)	200.8	659,041
2 Malignant neoplasms (cancer)	182.7	599,601
3 Accidents (unintentional injury)	52.7	173,040
4 Chronic lower respiratory diseases	47.8	156,979
5 Cerebrovascular diseases (stroke)	45.7	150,005
6 Alzheimer's disease	37.0	121,499
7 Diabetes mellitus (diabetes)	26.7	87,647
8 Nephritis, nephrosis (kidney disease)	15.7	51,565
9 Influenza & pneumonia	15.2	49,783
10 Suicide [Intentional Self-Harm]	**14.5**	**47,511**
11 Chronic liver disease and cirrhosis	13.5	44,358
12 Septicemia	11.7	38,431
13 Essential hypertension and renal disease	11.1	36,524
14 Parkinson's disease	10.8	35,311
15 Pneumonitis due to solids and liquids	5.8	19,184
- All other causes (Residual; > 15)	178.0	584,359
16 Homicide	5.8	19,141

- Older adults made up 16.5% of 2019 population, but 19.3% of suicides • Young made up 13.0% of 2019 population and 12.5% of suicides •
- Middle Aged made up 25.4% of the 2019 population, but were 34.2% of suicides •

1,358,796* Years of Potential Life Lost Before Age 75 (43,201 of 47,511 suicides are below age 75)

* alternate YPLL figure: 1,356,925 using individual years in calculations rather than 10-year age groups as above.

Many figures appearing here are derived or calculated from data in the following *official data sources:* downloaded 23 December 2020 from CDC's WONDER website: https://wonder.cdc.gov. • Other references cited on this page are listed on the State Data Page. •

suicide rate = (number of suicides by group / population of group) X 100,000

Suicide Data Page: 2019

Prepared for *AAS* by Christopher W. Drapeau, Ph.D. & John L. McIntosh, Ph.D.

23 December 2020

§ Alternate terms = **Survivors of Suicide Attempts** or those with **Lived Experience** (of suicide attempt)

Rate, Number, and Ranking of Suicide for Each U.S.A. State, 2019*

Rank	State [Division / Region]	Deaths	Rate
1	Wyoming [M / West]	170	29.4
2	Alaska [P / West]	210	28.7
3	Montana [M / West]	289	27.0
4	New Mexico [M / West]	513	24.5
5	Colorado [M / West]	1,312	22.8
6	Oregon [P / West]	906	21.5
7	Nevada [M / West]	642	20.8
8	Oklahoma [WSC / South]	816	20.6
8	South Dakota [WNC / Midwest]	182	20.6
10	Maine [NE / Northeast]	276	20.5
11	Idaho [M / West]	365	20.4
11	Utah [M / West]	654	20.4
13	Arizona [M / West]	1,419	19.5
14	New Hampshire [NE / Northeast]	255	18.8
15	Missouri [WNC / Midwest]	1,141	18.6
16	West Virginia [SA / South]	330	18.4
17	Arkansas [WSC / South]	548	18.2
18	Kansas [WNC / Midwest]	523	18.0
19	North Dakota [WNC / Midwest]	136	17.8
19	Tennessee [ESC / South]	1,219	17.8
21	Vermont [NE / Northeast]	110	17.6
22	Kentucky [ESC / South]	756	16.9
23	Iowa [WNC / Midwest]	528	16.7
24	Washington [P / West]	1,263	16.6
25	South Carolina [SA / South]	852	16.5
26	Alabama [ESC / South]	804	16.4
27	Florida [SA / South]	3,465	16.1
28	Nebraska [WNC / Midwest]	309	16.0
29	Hawaii [P / West]	224	15.8
30	Ohio [ENC / Midwest]	1,806	15.5
31	Louisiana [WSC / South]	704	15.1
32	Georgia [SA / South]	1,585	14.9
33	Pennsylvania [MA / Northeast]	1,896	14.8
34	Michigan [ENC / Midwest]	1,472	14.7
34	Minnesota [WNC / Midwest]	830	14.7
36	Mississippi [ESC / South]	436	14.6
37	Wisconsin [ENC / Midwest]	845	14.5
Nation		**47,511**	**14.5**
38	Indiana [ENC / Midwest]	972	14.4
39	Texas [WSC / South]	3,891	13.4
39	Virginia [SA / South]	1,140	13.4
41	North Carolina [SA / South]	1,358	12.9
42	Connecticut [NE / Northeast]	435	12.2
43	Rhode Island [NE / Northeast]	123	11.6
44	Delaware [SA / South]	111	11.4
44	Illinois [ENC / Midwest]	1,439	11.4
46	California [P / West]	4,436	11.2
47	Maryland [SA / South]	657	10.9
48	Massachusetts [NE / Northeast]	647	9.4
49	New York [MA / Northeast]	1,705	8.8
50	New Jersey [MA / Northeast]	762	8.6
51	District of Columbia [SA / South]	44	6.2

Caution: Annual fluctuations in state levels combined with often relatively small populations can make these data highly variable. The use of several years' data is preferable to conclusions based on single years alone.

Suggested citation: Drapeau, C. W., & McIntosh, J. L. (for the American Association of Suicidology). (2020). *U.S.A. suicide: 2019 Official final data.* Washington, DC: American Association of Suicidology, dated December 23, 2020, downloaded from http://www.suicidology.org.

Division [Abbreviation]	Rate	Number
Mountain [M]	21.6	5,364
West North Central [WNC]	17.0	3,649
East South Central [ESC]	16.8	3,215
West South Central [WSC]	14.7	5,959
South Atlantic [SA]	14.5	9,542
Nation	**14.5**	**47,511**
East North Central [ENC]	13.9	6,534
Pacific [P]	13.2	7,039
New England [NE]	12.4	1,846
Middle Atlantic [MA]	10.6	4,363

Region [Subdivision Abbreviations]	Rate	Number
West (M, P)	15.8	12,403
Midwest (WNC, ENC)	14.9	10,183
South (ESC, WSC, SA)	14.9	18,716
Nation	**14.5**	**47,511**
Northeast (NE, MA)	11.1	6,209

Source: Obtained 23 December 2020 from CDC/NCHS's *WONDER* (to appear in *Deaths: Final Data for 2019*, forthcoming) http://www.cdc.gov/nchs/products/nvsr.htm

[data are by place of residence]
[Suicide = ICD-10 Codes X60-X84, Y87.0, U03]

Note: All rates are per 100,000 population.

* Including the District of Columbia.

Suicide State Data Page: 2019
23 December 2020

Prepared by Christopher W. Drapeau, Ph.D. and John L. McIntosh, Ph.D. for

American Association of Suicidology
5221 Wisconsin Avenue, N.W.
Washington, DC 20015
(202) 237-2280

"to understand and prevent suicide as a means of promoting human well-being"

Visit the AAS website at:

http://www.suicidology.org

For other suicide data, and an archive of state data, visit the website below and click on the dropdown "Suicide Stats" menu:

https://jmcintos.pages.iu.edu

References from previous page

SAMHSA 2019 study (2020): Substance Abuse and Mental Health Services Administration. (2020). *Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health* (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved December 23, 2020 from https://www.samhsa.gov/data/report/2019/9-nsduh-annual-national-report

† Cerel, J., McIntosh, J. L., Neimeyer, R. A., Maple, M., & Marshall, D. (2014). The continuum of "survivorship": Definitional issues in the aftermath of suicide. *Suicide & Life-Threatening Behavior, 44*(6), 591-600.

Cerel, J. (2015, April 18). *We are all connected in suicidology: The continuum of "survivorship."* Plenary presentation at the 48th annual conference of the American Association of Suicidology, Atlanta, GA. [data from Cerel, Brown, Maple, Bush, van de Venne, Moore, & Flaherty, in progress; personal communication 20 Dec 2015]

Feigelman, W. Cerel, J., McIntosh, J. L., Brent, D., & Gutin, N. (2017). Suicide exposures and bereavement among American adults: Evidence from the 2016 General Social Survey. *Journal of Affective Disorders, 227*, 1-6. doi:10.1016/j.jad.2017.09.056

Jordan, J. R., & McIntosh, J. L. (Eds.). (2011). *Grief after suicide: Understanding the consequences and caring for the survivors.* New York: Routledge.

#vitalsigns
JUN 2018

CDC Vitalsigns™

Suicide rising across the US

45K Nearly 45,000 lives lost to suicide in 2016.

↑30% Suicide rates went up more than 30% in half of states since 1999.

54% More than half of people who died by suicide did not have a known mental health condition.

More than a mental health concern

Suicide is a leading cause of death in the US. Suicide rates increased in nearly every state from 1999 through 2016. Mental health conditions are often seen as the cause of suicide, but suicide is rarely caused by any single factor. In fact, many people who die by suicide are not known to have a diagnosed mental health condition at the time of death. Other problems often contribute to suicide, such as those related to relationships, substance use, physical health, and job, money, legal, or housing stress. Making sure government, public health, healthcare, employers, education, the media and community organizations are working together is important for preventing suicide. Public health departments can bring together these partners to focus on comprehensive state and community efforts with the greatest likelihood of preventing suicide.

States and communities can

- Identify and support people at risk of suicide.
- Teach coping and problem-solving skills to help people manage challenges with their relationships, jobs, health, or other concerns.
- Promote safe and supportive environments. This includes safely storing medications and firearms to reduce access among people at risk.
- Offer activities that bring people together so they feel connected and not alone.
- Connect people at risk to effective and coordinated mental and physical healthcare.
- Expand options for temporary help for those struggling to make ends meet.
- Prevent future risk of suicide among those who have lost a friend or loved one to suicide.

Want to learn more?
Visit: www.cdc.gov/vitalsigns

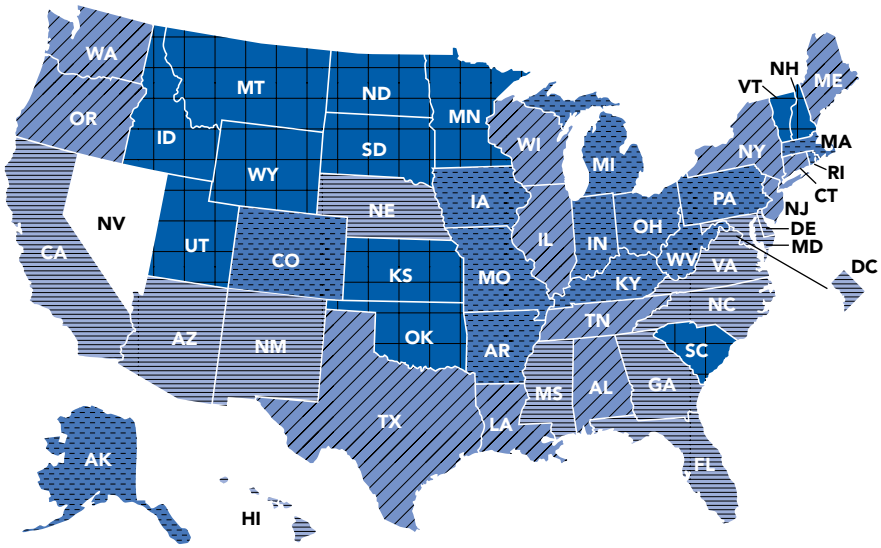
Centers for Disease Control and Prevention
National Center for Injury Prevention and Control

PROBLEM:

Suicide rates increased in almost every state.

Suicide rates rose across the US from 1999 to 2016.

Increase	38 - 58%
Increase	31 - 37%
Increase	19 - 30%
Increase	6 - 18%
Decrease	1%



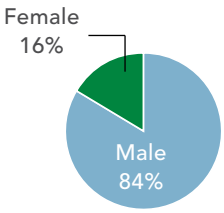
SOURCE: *CDC's National Vital Statistics System; CDC Vital Signs, June 2018.*

Differences exist among those with and without mental health conditions.

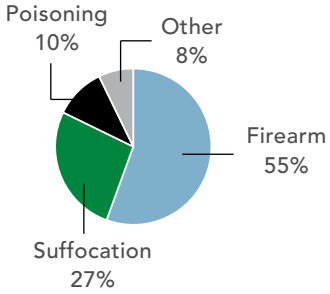
People without known mental health conditions were more likely to be male and to die by firearm.

No known mental health conditions

Sex

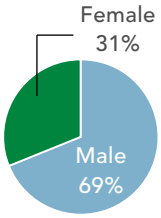


Method

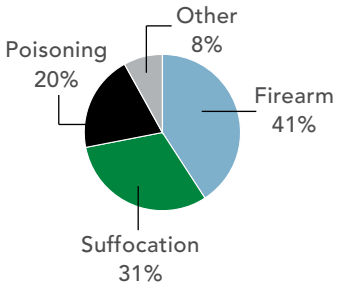


Known mental health conditions

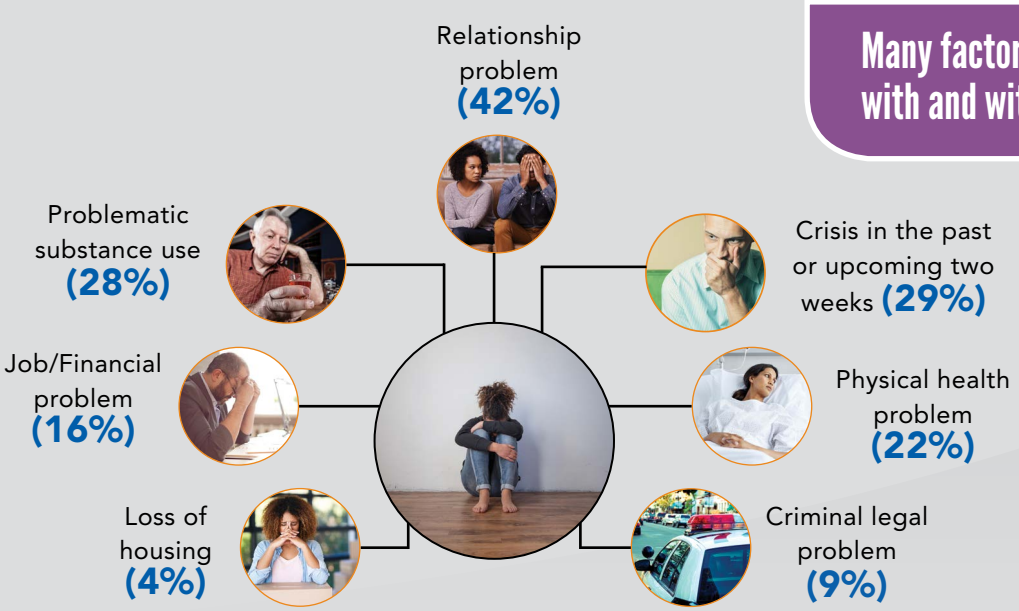
Sex



Method



Many factors contribute to suicide among those with and without known mental health conditions.



Note: Persons who died by suicide may have had multiple circumstances. Data on mental health conditions and other factors are from coroner/medical examiner and law enforcement reports. It is possible that mental health conditions or other circumstances could have been present and not diagnosed, known, or reported.

SOURCE: *CDC's National Violent Death Reporting System, data from 27 states participating in 2015.*

WHAT CAN WE DO TO PREVENT SUICIDE?

Preventing Suicide: A Technical Package of Policy, Programs, and Practices
https://go.usa.gov/xQBGc

Preventing suicide involves everyone in the community.

Provide financial support to individuals in need.

States can help ease unemployment and housing stress by providing temporary help.

Strengthen access to and delivery of care.

Health care systems can offer treatment options by phone or online where services are not widely available.

Create protective environments.

Employers can apply policies that create a healthy environment and reduce stigma about seeking help.

Connect people within their communities.

Communities can offer programs and events to increase a sense of belonging among residents.

Teach coping and problem-solving skills.

Schools can teach students skills to manage challenges like relationship and school problems.

Prevent future risk.

Media can describe helping resources and avoid headlines or details that increase risk.

Identify and support people at risk.

Everyone can learn the warning signs for suicide, how to respond, and where to get help.

Know the Suicide WARNING SIGNS

- Feeling like a burden
- Being isolated

- Increased anxiety
- Feeling trapped or in unbearable pain

- Increased substance use
- Looking for a way to access lethal means

- Increased anger or rage
- Extreme mood swings

- Expressing hopelessness
- Sleeping too little or too much

- Talking or posting about wanting to die
- Making plans for suicide

5 STEPS TO HELP SOMEONE AT RISK

1. Ask.
2. Keep them safe.
3. Be there.
4. Help them connect.
5. Follow up.

Find out why this can save a life by visiting:
www.BeThe1To.com

SOURCE: www.*BeThe1To.com*

WHAT CAN BE DONE

THE FEDERAL GOVERNMENT IS

- Tracking the problem to understand trends and the groups at greatest risk (for example, see www.cdc.gov/violenceprevention/nvdrs).
- Developing, implementing, and evaluating suicide prevention strategies.
- Supporting local, state, tribal, national, and other partners to prevent suicide (for example, see https://go.usa.gov/xQBGc).

STATES AND COMMUNITIES CAN

- Identify and support people at risk of suicide.
- Teach coping and problem-solving skills to help people manage challenges with relationships, jobs, health, or other concerns.
- Promote safe and supportive environments. This includes safely storing medications and firearms to reduce access among people at risk.
- Offer activities that bring people together so they feel connected and not alone.
- Connect people at risk to effective and coordinated mental and physical healthcare.
- Expand options for temporary assistance for those struggling to make ends meet.
- Prevent future risk of suicide among those who have lost a friend or loved one to suicide.

HEALTH CARE SYSTEMS CAN

- Provide high-quality, ongoing care focused on patient safety and suicide prevention.
- Make sure affordable and effective mental and physical healthcare is available where people live.
- Train providers in adopting proven treatments for patients at risk of suicide.

EMPLOYERS CAN

- Promote employee health and well-being, support employees at risk, and have plans in place to respond to people showing warning signs.
- Encourage employees to seek help, and provide referrals to mental health, substance use, legal, or financial counseling services as needed.

EVERYONE CAN

- Ask someone you are worried about if they're thinking about suicide.
- Keep them safe. Reduce access to lethal means for those at risk.
- Be there with them. Listen to what they need.
- Help them connect with ongoing support. You can start with the Lifeline (1-800-273-8255).
- Follow up to see how they're doing.
- Find out why this can save a life by visiting: www.BeThe1To.com.

The media can avoid increasing suicide risk (e.g., by not using dramatic headlines or providing explicit details) and encourage people to seek help.
View recommendations at:
www.ReportingOnSuicide.org

If you need help for yourself or someone else, please contact the
National Suicide Prevention Lifeline
Talk: **1-800-273-TALK (8255)**
Chat: **www.suicidepreventionlifeline.org**

www.cdc.gov/vitalsigns/suicide
www.cdc.gov/mmwr

For more information, please contact
Telephone: 1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 | Web: www.cdc.gov
Centers for Disease Control and Prevention
1600 Clifton Road NE, Atlanta, GA 30333
Publication date: June 7, 2018
CS292322-A

Table 2. Multivariate Associations of National Comorbidity Survey/*DSM-III-R* Disorders With Subsequent First Onset of Attempted Suicide in the Total Sample and Disaggregated Through Pathways Involving Onset of Ideation, Plans, Impulsive Attempts, and Planned Attempts*

	Total Sample, OR (95% CI)		OR (95% CI)		
	Attempt (n = 272)	Ideation (n = 795)	Plan Among Ideators (n = 230)	Impulsive Attempt Among Ideators Without a Plan (n = 145)	Planned Attempt Among Ideators With a Plan (n = 127)
Mood disorders					
Major depressive episode	11.0 (7.1-20.3)†	9.6 (7.5-12.3)†	1.7 (1.1-2.5)†	1.9 (1.3-2.9)†	2.1 (1.2-3.7)†
Dysthymia	7.8 (4.6-13.5)†	7.7 (5.9-10.1)†	1.9 (1.4-2.8)†	1.5 (0.9-2.5)	1.8 (1.1-2.9)†
Mania	29.7 (11.7-75.1)†	15.5 (8.9-26.8)†	4.0 (1.4-11.7)†	9.1 (1.7-50.3)†	3.2 (1.1-9.0)†
Any mood disorder	12.9 (7.8-21.3)†	10.7 (8.4-13.5)†	1.9 (1.3-2.8)†	1.7 (1.2-2.6)†	2.0 (1.2-3.4)†
Anxiety disorders					
Generalized anxiety disorder	5.6 (2.8-11.1)†	4.2 (3.1-5.9)†	2.6 (1.3-5.0)†	2.5 (1.2-5.5)†	1.1 (0.4-2.8)
Agoraphobia	2.8 (1.6-5.1)†	2.9 (2.2-3.9)†	1.5 (0.8-2.7)	1.1 (0.7-1.9)	1.2 (0.7-1.9)
Simple phobia	3.1 (1.8-5.3)†	2.9 (2.3-3.6)†	1.5 (1.1-2.2)†	1.2 (0.8-1.7)	1.3 (0.6-2.1)
Social phobia	2.1 (1.3-3.6)†	2.2 (1.7-2.9)†	1.5 (1.0-2.4)†	1.1 (0.9-1.4)	0.9 (0.6-1.3)
Panic disorder	5.6 (2.9-10.7)†	3.9 (2.8-5.3)†	1.6 (0.8-3.2)	2.0 (1.1-3.5)†	2.0 (1.0-4.0)†
Posttraumatic stress disorder	6.0 (3.4-10.7)†	5.1 (3.9-6.8)†	2.4 (1.7-3.3)†	1.7 (1.1-2.7)†	1.0 (0.6-1.6)
Any anxiety	3.2 (2.0-5.2)†	2.8 (2.2-3.5)†	1.7 (1.1-2.5)†	1.3 (1.0-1.7)†	1.0 (0.7-1.5)
Substance disorders					
Alcohol abuse	4.8 (2.8-8.1)†	3.4 (2.7-4.3)†	1.6 (1.1-2.4)†	1.5 (0.9-2.3)	1.5 (0.9-2.6)
Alcohol dependence	6.5 (3.6-11.5)†	4.6 (3.5-6.1)†	1.9 (1.3-2.9)†	1.4 (0.8-2.5)	1.8 (0.9-3.7)
Drug abuse	5.9 (3.4-10.2)†	4.9 (3.8-6.1)†	1.6 (1.1-2.3)†	1.5 (1.0-2.4)†	1.3 (1.0-1.9)†
Drug dependence	5.8 (3.3-10.1)†	5.3 (4.1-6.9)†	1.5 (1.0-2.3)†	1.4 (0.8-2.3)	1.3 (0.8-2.1)
Any substance disorder	5.8 (3.5-9.7)†	3.9 (3.1-4.9)†	1.8 (1.3-2.6)†	1.8 (1.3-2.7)†	1.4 (0.9-1.9)
Other disorders					
Conduct disorder	4.2 (2.5-6.9)†	3.2 (2.6-3.9)†	2.1 (1.5-2.9)†	1.6 (1.0-2.5)†	1.2 (0.8-1.8)
Adult antisocial behavior	5.7 (3.1-10.3)†	4.2 (3.2-5.6)†	2.1 (1.2-3.6)†	2.9 (1.5-5.5)†	0.7 (0.4-1.4)
Antisocial personality disorder	5.7 (2.9-11.2)†	4.6 (3.2-6.5)†	2.2 (1.2-4.0)†	1.8 (0.8-4.2)	1.1 (0.5-2.2)
Nonaffective psychosis	5.7 (2.6-12.4)†	4.2 (2.7-6.7)†	2.4 (1.0-6.1)†	2.8 (0.9-8.8)	0.8 (0.2-3.7)
No. of total disorders					
Any	6.7 (4.1-11.0)†	5.7 (4.4-7.3)†	1.4 (0.9-2.3)	1.7 (1.1-2.5)†	1.0 (0.6-1.7)
1	3.8 (2.1-6.9)†	3.4 (2.7-4.3)†	1.2 (0.7-2.1)	1.6 (1.0-2.6)†	0.5 (0.2-1.0)
2	6.1 (4.1-9.2)†	6.1 (4.5-8.3)†	1.0 (0.2-5.3)	1.4 (0.8-2.6)	1.0 (0.5-2.1)
≥3	19.7 (13.2-29.3)†	14.3 (11.4-18.1)†	2.4 (0.9-6.4)	2.3 (1.4-3.7)†	1.1 (0.6-2.1)

**Odds ratios (ORs) were obtained by exponentiating coefficients from discrete-time survival models. Disorders were defined without diagnostic hierarchy rules. The 95% confidence intervals (CIs) were obtained using the method of Jacknife Repeated Replications to adjust for clustering and weighting of data. Each column in this table presents the results of 22 models. Each model controls for person-year and the sociodemographic variables in Table 1. In addition, each model contains (a) exactly 1 of the 17 individual disorders, (b) exactly 1 of the 3 summary measures (any mood, any anxiety, any substance), (c) a single dichotomous measure that distinguishes between respondents with any disorders and those with no disorders, or (d) the set of summary measures of number of disorders (exactly 1, exactly 2, and 3 or more).*

†P<.05 by 2-sided test.

Inspection of the distribution of the summary risk factor count in the sample of person-years used in the survival analysis (**Table 4**) shows that the extremely high OR of a suicide attempt associated with having 5 or more risk factors accounts for only 7.0% of the people who made a lifetime attempt. This is because only a tiny fraction of the population (0.4%) had this large number of risk factors. People with 3 or more risk factors, who made up 9.2% of the population, accounted for 55.1% of all people who made a lifetime suicide attempt.

COMMENT

The results reported here are limited by the fact that the NCS is a cross-sectional survey in which information about lifetime suicide behaviors is based on retrospective reports. Because of this limitation, prevalences are likely to be lower bound estimates. The estimated effects of cohort and other risk factors could be caused, at least in part, by systematic differences in accuracy of recall related to these risk factors.[36] Furthermore, no reliability or validity data were obtained on the measures of ideation, plans, attempts, or lethality. Finally, the ability to study the distribution and predictors of lethality is constrained by the fact that the sample excluded people who completed suicide. This selection bias might account for the failure to find a significant positive association between age of first attempt and lethality of intent, an association that has been found when both attempters who died and those who survived have been considered in the same analysis.[37]

Within the context of these limitations, the 4.6% estimated lifetime prevalence of attempted suicide is above the high end of the range of estimates reported in previous US general population surveys.[15] Also, the 13.5% estimated prevalence of suicide ideation is at the high end of the range of previous estimates.[19,38,39] However, the question wording to assess ideation varied in important ways in earlier surveys,[40] sometimes asking about "thoughts" of suicide,[19] "serious" thoughts,[39] and "serious thoughts

Downloaded from www.archgenpsychiatry.com at ARIZ HEALTH SCIENCES-LIBRARY, on October 31, 2005
©1999 American Medical Association. All rights reserved.

VIEWPOINT

Self-injury Is the Eighth Leading Cause of Death in the United States

It Is Time to Pay Attention

Ian R. H. Rockett, PhD, MPH
Department of Epidemiology, School of Public Health and the Injury Control Research Center, West Virginia University, Morgantown.

Eric D. Caine, MD
Department of Psychiatry and the Injury Control Research Center for Suicide Prevention, University of Rochester Medical Center, Rochester, New York.

Corresponding Author: Ian R. H. Rockett, PhD, MPH, Department of Epidemiology, School of Public Health, West Virginia University, PO Box 9190, Morgantown, WV 26506-9190 (irockett@hsc.wvu.edu).

Establishing a person's intention to die has been a central element separating suicides from fatal self-injurious acts that are labeled "accidents" or "unintentional" deaths. We argue that this is a false dichotomy—certainly at the level of populations—that masks the overall magnitude of fatalities arising from deliberate, self-destructive behaviors. In so doing, it mutes the urgency for demanding effective preventive interventions and is particularly problematic as the nation experiences a persisting and growing epidemic of opioid and other drug-poisoning deaths.[1] Firearm trauma and hanging/asphyxiation, the 2 leading methods of suicide, typically generate ample forensic evidence for assuring accurate determinations by medical examiners and coroners. However, corroborative evidence is less available for poisoning, the third leading method of suicide overall, and first among women. Parenthetically, we acknowledge that the Centers for Disease Control and Prevention use "unintentional injury" in lieu of the term *accident* for surveillance and prevention purposes. However, medical examiners and coroners remain bound by statutes in using "accident" as 1 of 6 manner-of-death entries (homicide, suicide, accident, undetermined, natural causes, and unknown) that alternatively appear on death certificates.

There are abundant data indicating the conjoined nature of the groups dying by suicide and "accident." Longitudinal cohort studies of survivors of self-harm show excess risk for both manners of death.[2] Moreover, accident survivors manifest an elevated risk for suicide, as do survivors of self-poisoning specifically.[3] Results from 2 recent overseas studies further reinforce the complexity of accurately distinguishing suicide from accident poisoning deaths. One suggested that as much as 43% of the sharp increase in South Korea's suicide rate was an artifact of more accurate determination that offset a decline in the proportion of accidental (predominantly poisoning) deaths.[4] The second, an in-depth, records-based study, detected an increasing trend of suicide undercounting in England, which the investigators attributed to misclassification of pharmaceutical drug-intoxication "accident" deaths.[5]

Classifying deaths arising from intoxication with medications or illicit drugs as "accidents," when the fundamental behaviors most often were intentional (irrespective of "suicidal intent"), serves as a barrier to prevention. To help circumvent these problems, we teamed up with colleagues to propose a new category, *death from drug self-intoxication* (DDSI).[6] Death from drug self-intoxication encompasses all drug-intoxication suicides and most accidental and undetermined drug-intoxication deaths and emphasizes that hazardous premorbid behaviors are deliberate—whether or not there is an explicit intention to die on the day of death. These self-determined behaviors profoundly alter the probability of adverse events, including death, just as hazardous or intoxicated driving increases the likelihood of motor vehicle traffic deaths (which no longer are called "accidents"). Operationalization of DDSI would enable suicide and substance abuse researchers and prevention scientists to end their dependence on the medicolegal determinations of manner of death, which vary according to statutory guidelines for the level of certainty required to determine suicide and to related information bias (ie, lack of proof-positive indication of intent), type of medical examiner or coroner system in each state or county, rigorousness with which cases are investigated, and the force of local considerations that diminish suicide detection. It also would open the door for researchers to examine the common risks that link or distinguish fatal drug intoxications.

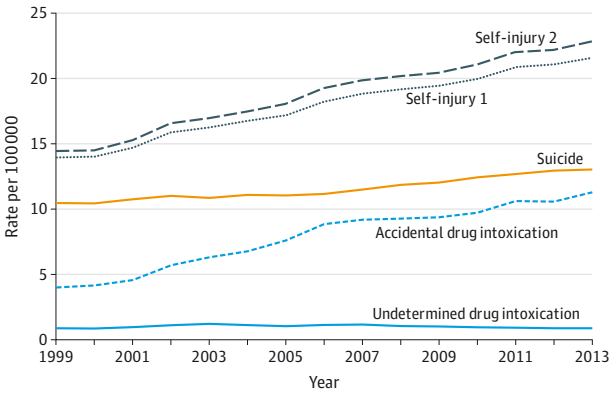
To more accurately assess the magnitude of self-inflicted injury deaths in the United States, we combined estimated nonsuicide DDSIs with total registered suicides to portray the trend as well as the magnitude of rates from 1999 to 2013 (**Figure**), using data from the Multiple Cause-of-Death public use files created by the National Center for Health Statistics. We computed 2 series of estimated self-injury mortality rates. Series 1 assumed that 70% of the drug-intoxication accident deaths and 80% of the undetermined drug-intoxication deaths, at ages 15 years and older, were DDSIs. Series 2 substituted corresponding constants of 80% and 90%. Whereas the suicide rate rose 24% over the observation period, our more conservative estimate of the self-injury mortality rate increased by 55% and our higher estimate by 58%.

At 68 298 or 72 137 self-injury deaths for 2013, the estimated counts from series 1 and 2 were, respectively, 66% and 75% higher than the suicide count of 41 149. Suicide alone is officially the 10th leading cause of death; either self-injury mortality estimate would clearly constitute the eighth leading cause, exceeding kidney disease (47 112) and pneumonia and influenza (56 979).[7] We recognize that assumptions underlying these estimates are simplifications. For example, we made no provision to include motor vehicle traffic deaths that may have

Copyright 2015 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Arizona Health Sciences Library User on 10/30/2015

Figure. Selected Manner and Cause-of-Death Rates per 100 000 Population: United States, 1999-2013



The year-specific self-injury 1 death rate equals total suicide rate, +0.7 and 0.8 of respective accidental and undetermined drug-intoxication death rates for ages 15 years and older. The corresponding self-injury 2 death rate substituted constants of 0.8 and 0.9.

been suicides or reflected intentional high-risk, hazardous driving.

We do not expect that the medicolegal manner-of-death components (ie, homicide, suicide, accident, undetermined, and natural causes) will be modified to accommodate a more nuanced classification of drug-intoxication deaths. More feasible, inclusion of a new subcategory on the death certificate for recording premorbid substance misuse and abuse would enhance the quality of data needed to discern fatal self-injurious behaviors. Corroborative evidence for justifying an affirmative entry could include needle marks on the corpse or documentation of physician or pharmacy shopping from prescription monitoring programs.

We offer a caution regarding substance use and abuse prevention. The contemporary focus on fatal "prescription drug overdoses" may be inadvertently skewing consideration toward one source of lethal compounds rather than capturing the necessary breadth of substances that characterize the fluid nature of drug abuse and misuse, where addicted individuals, and those who are experimenting with opiates and other agents, shift their demand to whatever drugs are accessible and cheaper, for example, away from prescribed oxycodone to injected or snorted heroin.

However one finally chooses to label drug self-intoxication fatalities more precisely, broad reliance on describing them as accidents (unintentional injury deaths) obscures the extraordinary social, economic, and health burden that is being generated by deliberate self-destructive behaviors that either are overtly intended to kill or are so hazardous they do frequently. The nation must recognize and acknowledge the plethora of premature injury deaths that reflect such self-harm and develop a sense of urgency matching that previously shown other seemingly insurmountable health crises. The year 1964 marked the release of the Surgeon General's inaugural report on *Smoking and Health*. Who at that time could have anticipated the radical transformation in the attitudes of physicians and the public toward cigarette smoking that has been crucial in preventing numerous deaths from cancers and vascular diseases? By our reckoning, the eighth leading cause of death warrants similar urgent attention.

ARTICLE INFORMATION

Published Online: September 16, 2015. doi:10.1001/jamapsychiatry.2015.1418.

Conflict of Interest Disclosures: None reported.

Funding/Support: This research received support from the Centers for Disease Control and Prevention (grants 5R49CE002109 and R49CE002093).

Role of the Funder/Sponsor: The Centers for Disease Control and Prevention had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Injury Prevention and Control, the Centers for Disease Control and Prevention.

Additional Contributions: We thank Randy L. Hanzlick, MD (Fulton County Medical Examiner's Center and Department of Pathology, Emory School of Medicine, Atlanta, Georgia), for valuable comments on the manuscript. He received no compensation for his contributions.

REFERENCES

1. Rockett IRH, Kapusta ND, Coben JH. Beyond suicide: action needed to improve self-injury mortality accounting. *JAMA Psychiatry*. 2014;71(3):231-232.

2. Bergen H, Hawton K, Waters K, et al. Premature death after self-harm: a multicentre cohort study. *Lancet*. 2012;380(9853):1568-1574.

3. Finkelstein Y, Macdonald EM, Hollands S, et al; Canadian Drug Safety and Effectiveness Research Network (CDSERN). Risk of suicide following deliberate self-poisoning. *JAMA Psychiatry*. 2015;72(6):570-575.

4. Chan CH, Caine ED, Chang SS, Lee WJ, Cha ES, Yip PS. The impact of improving suicide death classification in South Korea: a comparison with Japan and Hong Kong. *PLoS One*. 2015;10(5):e0125730.

5. Gunnell D, Bennewith O, Simkin S, et al. Time trends in coroners' use of different verdicts for possible suicides and their impact on officially reported incidence of suicide in England: 1990-2005. *Psychol Med*. 2013;43(7):1415-1422.

6. Rockett IRH, Smith GS, Caine ED, et al. Confronting death from drug self-intoxication (DDSI): prevention through a better definition. *Am J Public Health*. 2014;104(12):e49-e55.

7. Kochanek KD, Murphy SL, Xu J, Arias E. Mortality in the United States, 2013. *NCHS Data Brief*. 2014;178(178):1-8.

Copyright 2015 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Arizona Health Sciences Library User on 10/30/2015

TABLE 1. Chi-Square Statistics for the Kruskal-Wallis ANOVA of Ranks for 954 Patients With Major Affective Disorder Who Did or Did Not Commit Suicide

Symptom	Chi-Square χ^2 (df=2)	p	ANOVA F (df=2, 951)[a]	p
Hopelessness	7.79	0.020	2.34	0.097
Alcohol abuse	5.73	0.057	2.43	0.089
Loss of interest or pleasure (anhedonia)	8.79	0.012	3.74	0.035
Psychic anxiety	6.36	0.042	3.27	0.038
Suicidal ideation	4.48	0.106	2.10	0.123
Suicide attempts	3.03	0.220	1.90	0.150
Obsessive-compulsive features	4.57	0.102	2.97	0.052
Indecisiveness	6.34	0.042	3.57	0.029
Diminished concentration	7.84	0.020	3.11	0.045
Global insomnia	6.58	0.037	2.39	0.096

[a]For suicidal ideation, df=2, 950.

TABLE 2. Probability Values for Mann-Whitney U Statistics Comparing 954 Patients With Affective Disorder Who Committed Suicide Within 1 Year (Short-Term) or 2-10 Years (Long-Term) and Patients Who Did Not Commit Suicide

Symptom	Short-Term Suicide p	Long-Term Suicide p
Hopelessness	0.463	0.007
Alcohol abuse	0.029	0.372
Loss of interest or pleasure (anhedonia)	0.005	0.223
Psychic anxiety	0.012	0.879
Suicidal ideation	0.613	0.041
Suicide attempts	0.815	0.086
Obsessive-compulsive features	0.063	0.303
Indecisiveness	0.085	0.062
Diminished concentration	0.028	0.078
Global insomnia	0.011	0.765

RESULTS

Thirty-two (3%) of the 954 patients had committed suicide. Thirteen (41%) of these suicides occurred during the first year of follow-up: three (9%) during the first 3 months and seven (22%) during the first 6 months. Nineteen (59%) of the suicides occurred during follow-up years 2–10.

Previously reported univariate analyses (10) showed that no specific RDC type or subtype of major affective disorder had a significantly higher incidence of suicide than any of the others.

not committed suicide are given in table 2. Symptoms that were significantly more severe among those who committed suicide within 13 months than among those who did not commit suicide were loss of interest or pleasure (anhedonia), psychic anxiety, obsessive-compulsive features, global insomnia, and alcohol abuse.

The symptom (not the disorder) of panic attacks was present at the intake SADS evaluation in eight (62%) of 13 patients who committed suicide within 1 year but only 262 (28%) of 922 patients who did not commit suicide and four (21%) of the 19 patients who committed suicide in 2–10 years. Despite the small number of suicides overall, this result cannot be attributed to chance alone (χ^2=7.43, df=2, p=0.024)

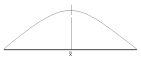
The unstructured intake interview

A. ID: Identifying data (age, sex, marital status, relevant special characteristics such as deafness, retardation, language barrier)
B. CC: Presenting or current complaints (verbatim first complaint, problems, stressors, symptoms, requests)
C. PRECIP: Precipitating event (what made the patient come in/call *TODAY?*);
D. HX: Relevant history and more detailed description of presenting problems (stressors, symptoms, recent changes)
E. DTS/O : Information about danger to self or others;
F. TX/PAST TX: Current mental health treatment (include names of provider); relevant past mental health treatment;
G. HEALTH: Currently relevant physical illnesses and injuries, and their treatment (inquire specifically about LOC, car wrecks)
H. MEDICATIONS: Current and Past
I. CD: Drug and alcohol use, abuse, and dependency (current and past);
J. LIFE: Current life situation (living arrangements, employment [current and past if relevant; commensurate with abilities/education?], family/marital activities, recreation/social support)
K. SS/LEGAL: Social service involvement/Legal system involvement: general assistance (welfare), food stamps, medicare; legal problems, criminal history, probation status
L. FAMILY HISTORY: Relevant and significant to current situation; can be more detailed if intake is for purpose of subsequent therapy (e.g., dynamic-oriented);
M. BEHAVIOR: Relevant behavior during interview (cooperation, appropriateness)
N. MS: Mental State (appearance, cooperation, orientation, mood, affect, unusual behavior, under the influence?, associations and thought processes, stream of speech, perceptual distortions, memory function, fund of information, judgment, insight, motivation for help/tx, self-esteem)
O. IMPRESSION: Conclusion: Diagnostic and otherwise
P. PLAN: Treatment plan (include necessary consultation, need for further information [e.g., r/o], referral, final disposition, follow-up instructions)

Measurement

I. Truly Basic Statistical Concepts:

A. For a distribution



1. $\overline{X} = \frac{\Sigma X}{N}$

2. $S_X = \sqrt{\frac{\Sigma X^2}{N} - (\frac{\Sigma X}{N})^2}$

To what extent is a sample statistic representative of a population?

3. $\sigma_{M_X} = \frac{\sigma_x}{\sqrt{N}}$

4. $\sigma_{S_X} = \frac{\sigma_x}{\sqrt{2N}}$

B. Pearson Correlation Coefficient (***r***)

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{(N\Sigma X^2 - (\Sigma X)^2)(N\Sigma Y^2 - (\Sigma Y)^2)}}$$

1. Long formula:

$$r_{xy} = \frac{\Sigma xy}{(N)(SD_x)(SD_y)}$$

2. Shorter formula:

$$r_{xy} = \frac{\Sigma z_x z_y}{N}$$

3. Conceptual formula:

4. $\sigma_r = \frac{1 - \rho_{xy}^2}{\sqrt{N}}$ Standard error of correlation

II. Considerations in generating test items

A. Select item type

B. What level of "difficulty"?

1. σ^2 of total scores should be maximized if goal is to provide rank-ordering of examinees
2. Items of medium difficulty tend to produce distributions with largest σ^2;

Item difficulty $p = \frac{R}{N}$, where $\boldsymbol{R}$ = # getting item "right" (endorsing in keyed direction) for dichotomously-scored items

C. Item Discrimination Statistic -- pearson correlation of item score to total score r_{it} should be > 0, preferably > .30

D. Predicting Mean Total Score $\bar{X} = n\bar{\rho}$

E. Predicting test σ :

1. σ . ¼ range
2. More precisely, $\sigma_T = n\sigma_{\bar{p}}\bar{r}_{it}$; $\sigma_{\bar{p}} = \sqrt{\bar{p}(1-\bar{p})}$;

F. Predicting Test Reliability (Kuder-Richardson formula 21)

1. For binary items: $r_{tt} = \frac{n(\sigma_t^2) - \bar{x}_t(n - \bar{x}_t)}{(n-1)\sigma_t^2}$

G. Finally, a few testing considerations ...

1. Speeded?
2. Write 2-3 times as many items as you wish to use in final version of test
3. # students in validation sample
4. Correcting for Guessing

$X_{corrected} = R - \frac{W}{(O-1)}$

$\boldsymbol{R}$ = # items respondent got correct

$\boldsymbol{W}$ = # items got wrong, not counting omits

$\boldsymbol{O}$ = # options (T-F, $\boldsymbol{O}$ = 2)

also can correct $\boldsymbol{p}$ for guessing $p_{corrected} = \frac{N_R - \frac{N_W}{(O-1)}}{N_{TOT}}$

N_R = # individuals getting item right

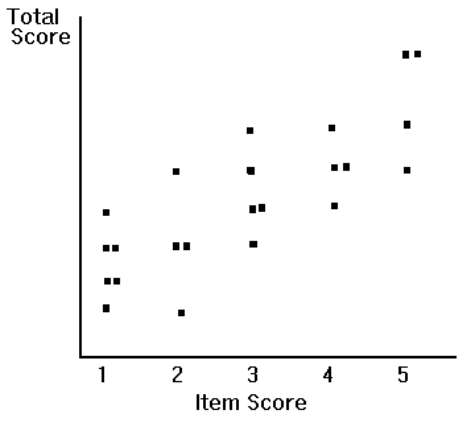
N_W = # individuals getting item wrong

III. Classical Item Analysis

A. Item variance-Covariance Matrix:

B. Discrimination statistic (r_{it}) -- pearson, point biserial, biserial correlations

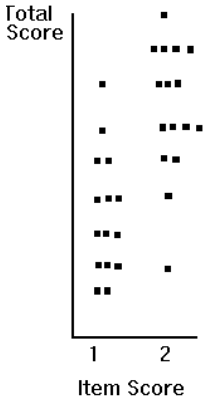
1. Likert Scale -- pearson ***r***



Good Item: pearson ***r*** high positive (.70) so that individuals high on trait also score high on total test.

$$r_{it} = \frac{N\Sigma XY_i - (\Sigma X)(\Sigma Y_i)}{\sqrt{(N\Sigma X^2 - (\Sigma X)^2)(N\Sigma Y_i^2 - (\Sigma Y_i)^2)}}$$

2. Dichotomous (T-F, or mult choice with right/wrong) -- point biserial correlation



$$r_{pbis} = \frac{M_i - M_X}{S_X}\sqrt{\frac{p_i}{1-p_i}}$$

Good Item (.75)

M_i = Mean score of those choosing item
M_x = Mean score on test

r_{pbis} is applicable only when one variable is binary

Correct Option r_{pbis}	Evaluation	Rationale
> .30	good	HI students choose, LO students avoid correct
0 - .29	weak	equally attractive to HI & LO students
< 0	horrid	HI students avoid, LO students choose correct

3. If item multiple choice, can also compute r_{pbis} for incorrect options; r_{pbis} should be negative

Incorrect Option r_{pbis}	Evaluation	Rationale
> 0	bad	HI students select, LO students avoid incorrect
< 0	good	HI students avoid, LO students select incorrect

a. Can also compute r_{pbis} for omits for a given item; r_{pbis} should be negative

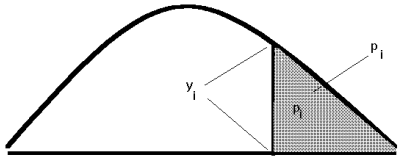
$$r_{pbis} = \frac{M_o - M_X}{S_X} \sqrt{\frac{p_o}{1 - p_o}}$$

Omit r_{pbis}	Evaluation	Rationale
> 0	bad	HI students omit, LO students respond to item
< 0	good	HI students respond, LO students omit item

b. (1) The biserial correlation

c. $r_{bis} = \frac{M_i - M_X}{S_X} \frac{p_i}{y_i}$

to determine y_i, use normal probability table



y_i is height (proportion) at this point on normal probability curve (e.g., @ p_i = .30, y_i = .35

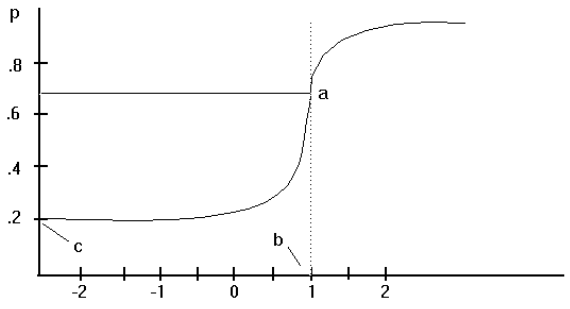
Relationship between point biserial and biserial:

$$r_{bis} = r_{pbis} \frac{\sqrt{p(1-p)}}{y}$$

latter term always > 1, therefore $r_{bis} > r_{pbis}$ (in absolute value)

IV. A brief version of Item Response Theory

A. Item Characteristic Curve (ICC)



f	1	6	2	3	2	3	3
R	0	0	0	0	1	2	3
p	0	0	0	0	.5	.67	1.0

ICC

θ = ability level
f = frequency of persons with total score θ
R = # persons with total score θ getting item right
p = probability of getting item right given total score θ

B. Three-parameter Logistic (*S*-shaped) model -- describes ICC

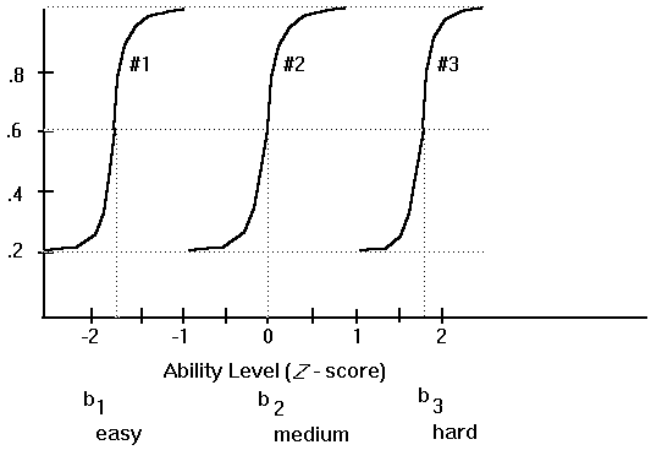
$$p(\theta) = c + (1-c)\frac{1}{1+e^{-a(\theta-b)}}$$

θ is latent construct of ability level

a, *b*, & *c* are constants
e . 2.71828 (ln(*e*) = 1)

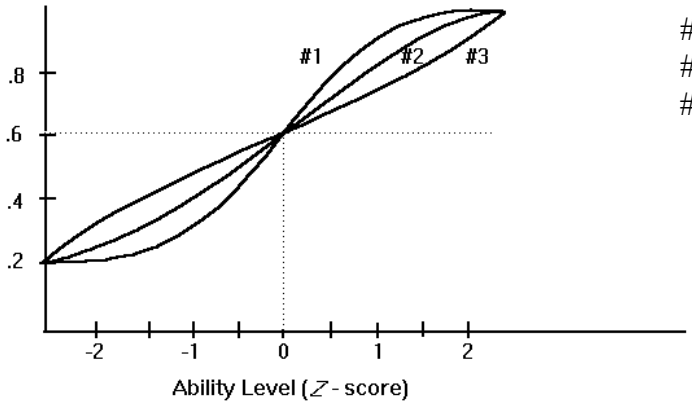
C. Utility of this model:

1. Difficulty parameter (*b*)



Three hypothetical items

2. Discrimination parameter (*a*), indicating discrimination power for examinees at ability level at *b*



#1, $a = 1.5$ highly discriminating
#2, $a = 1.0$ moderately discrimination
#3, $a = 0.5$ low discrimination

D. Other models (including 3 parameter model):

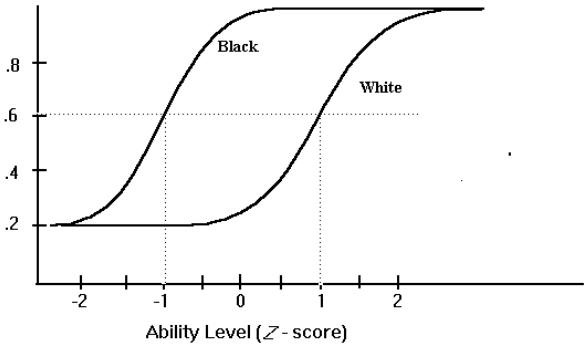
1. $p(\theta) = c + (1-c)\dfrac{1}{1+e^{-a(\theta-b)}}$

2. $p(\theta) = \dfrac{1}{1+e^{-a(\theta-b)}}$ assume $c = 0$, or $c = k$

3. $p(\theta) = \dfrac{1}{1+e^{-(\theta-b)}}$ assume $c = 0$, $a = 1$

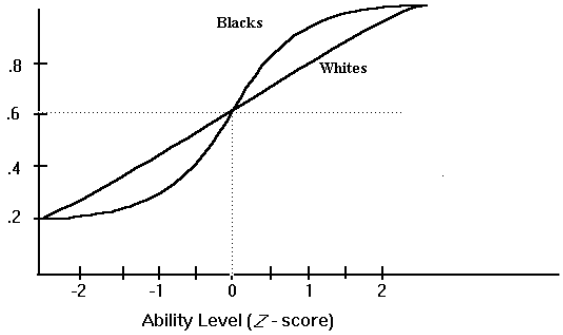
E. Other applications of item response theory -- racial bias in testing

1.



Item is more difficult for whites than for blacks at *all* ability levels $b_w \neq b_b$

2.



also an indication of a biased item $a_w \neq a_b$

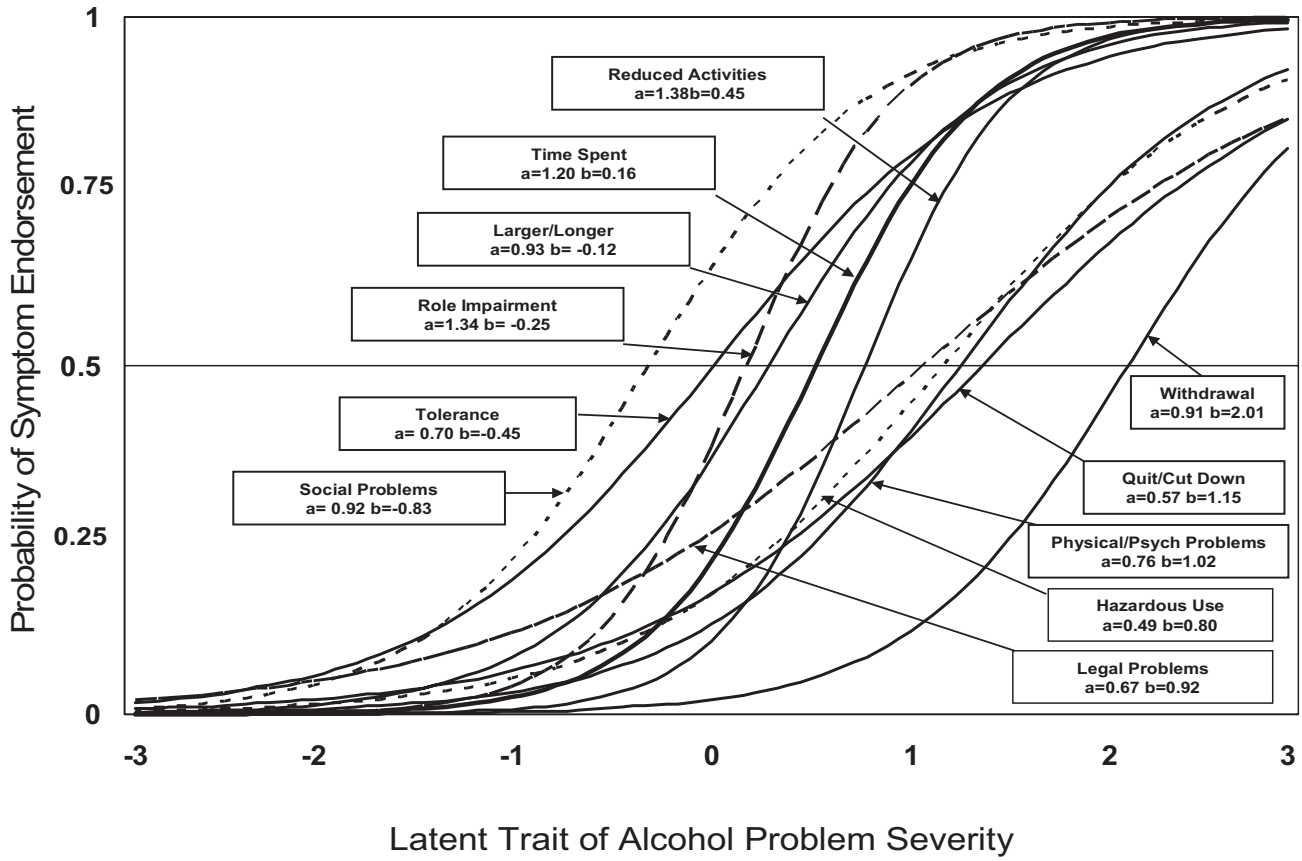


Figure 1. Item response curves (IRCs) for the 11 *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (*DSM–IV*), alcohol use disorder criteria. IRCs illustrate the probability of symptom endorsement (*y*-axis) across a latent trait of alcohol problem severity (*x*-axis). Item threshold (shown numerically as the "b" parameter) is illustrated by the point on the latent trait at which the probability of symptom endorsement is 50%; higher thresholds indicate greater severity. Item discrimination (shown numerically as the "a" parameter) is illustrated by the slope of an IRC at its threshold value; higher numbers and steeper slopes indicate better discrimination. IRCs for *DSM–IV* dependence symptoms have solid lines; IRCs for *DSM–IV* abuse symptoms have dashed lines.

clearly showed a range of discrimination and threshold values, the figures indicate that for both alcohol and cannabis, IRCs did not clearly distinguish between the *DSM–IV* abuse and dependence symptoms. Instead, for both alcohol and cannabis, there was a wide range of threshold values within each symptom group. Abuse and dependence symptoms were mixed in threshold and showed no clear pattern of identifying less severe and more severe symptom groups. In the alcohol data (see Figure 1), the abuse symptom of social problems had the lowest threshold value and relatively high discrimination. Moving progressively higher on the severity trait were the thresholds for tolerance, role impairment, larger/longer, and time spent using, followed by reduced activities. Role impairment, time spent using, and reduced activities had the highest discrimination values of the AUD symptoms, whereas discrimination was relatively low for tolerance. Next, there were four symptoms with higher threshold values and low discrimination, which also appear to provide fairly redundant psychometric information as indicated by their densely clustered IRCs: the abuse symptoms of hazardous use and legal problems, and the dependence symptoms of quit/cut down and psychological–physical problems. Alcohol withdrawal had the highest threshold value, but its discrimination was moderate.

A comparison of Figures 1 and 2 indicates far more similarities than differences between alcohol and cannabis in terms of the performance of diagnostic criteria. For cannabis, role impairment showed the lowest threshold value, followed by time spent using. Both of these items showed high discrimination. The next lowest threshold values were for social problems, tolerance, and larger/longer, all of which showed moderate discrimination, followed by reduced activities, which had high discrimination. The lowest discrimination values for cannabis were the CUD symptoms with the highest thresholds: hazardous use, legal problems, quit/cut down, and physical–psychological problems. Cannabis symptoms of hazardous use and legal problems had similar IRCs, indicating that they provide largely redundant psychometric information.

Gender Differences

We used differential item functioning (DIF) analyses to test for gender differences in item thresholds while controlling for overall

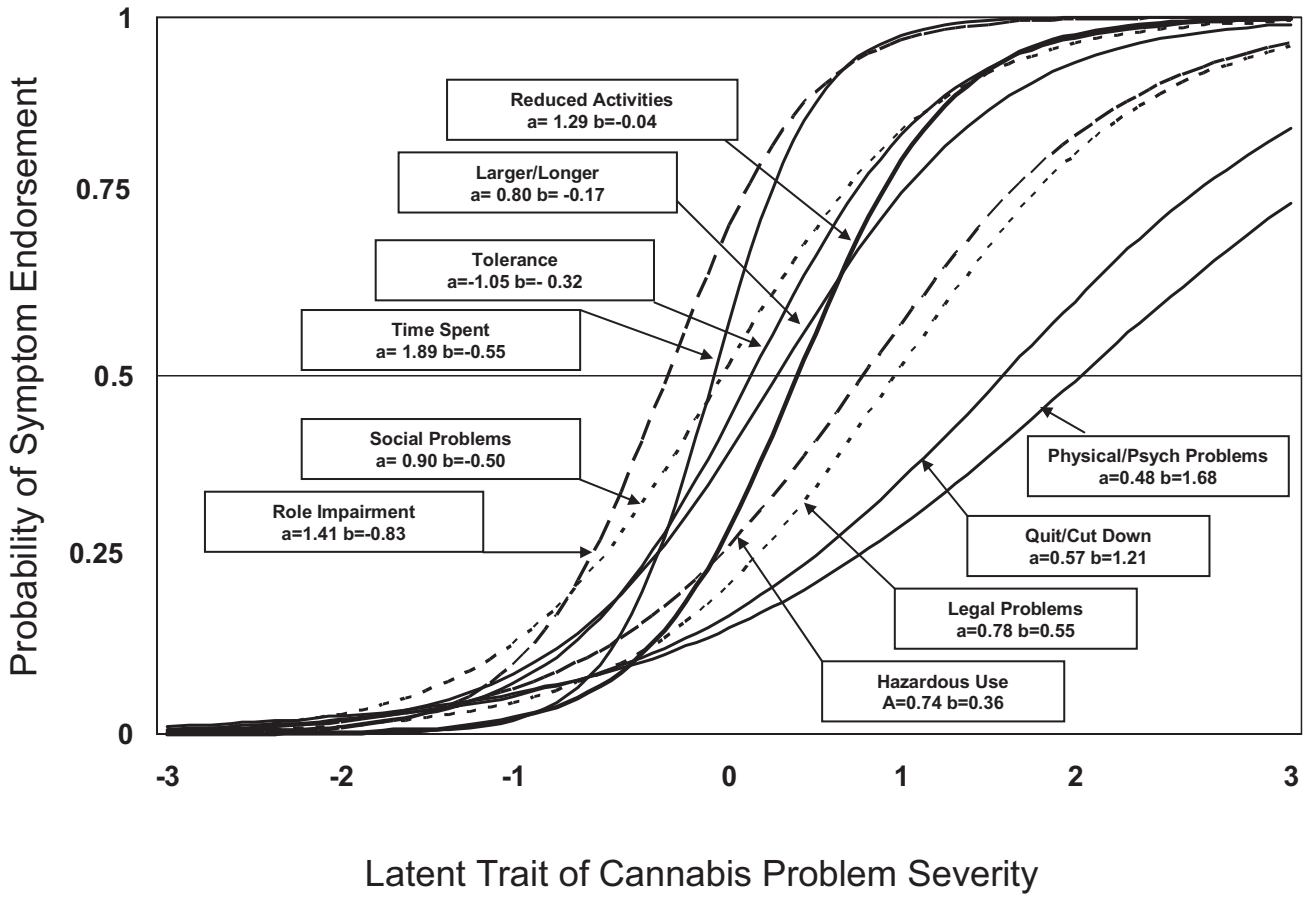


Figure 2. Item response curves (IRCs) for the 10 *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (*DSM–IV*), cannabis use disorder criteria. IRCs illustrate the probability of symptom endorsement (*y*-axis) across a latent trait of alcohol problem severity (*x*-axis). Item threshold (shown numerically as the "b" parameter) is illustrated by the point on the latent trait at which the probability of symptom endorsement is 50%; higher thresholds indicate greater severity. Item discrimination (shown numerically as the "a" parameter) is illustrated by the slope of an IRC at its threshold value; higher numbers and steeper slopes indicate better discrimination. IRCs for *DSM–IV* dependence symptoms have solid lines; IRCs for *DSM–IV* abuse symptoms have dashed lines.

substance problem severity, using the Mantel-Haenszel (MH) odds ratio statistic; p values were set to .01 to guard against Type I error. DIF was examined for alcohol symptoms (293 males and 171 females) and cannabis symptoms (272 males and 145 females). Gender differences were found for 4 of the 11 AUD symptoms. Controlling for overall alcohol problem severity, we found that females were less likely to be assigned the abuse symptoms of hazardous use (MH odds ratio = 0.29, $p < .001$) and legal problems (MH odds ratio = 0.20, $p < .001$). These results indicate that females tend to exhibit these symptoms at higher levels of alcohol problem severity than do males. In contrast, after controlling for overall severity, we found that females were more likely to have the dependence symptoms of reduced activities (MH odds ratio = 2.6, $p = .001$) and physical–psychological problems (MH odds ratio = 2.9, $p < .001$). These results suggest that females tend to show these symptoms at lower levels of alcohol problem severity compared with males.

Gender differences were obtained for 3 of the 10 CUD symptoms. Our findings were similar to the results for alcohol; after controlling for overall cannabis problem severity, we found that females were less likely to be assigned the cannabis abuse symptoms of hazardous use (MH odds ratio = 0.44, $p < .005$) and legal problems (MH odds ratio = 0.26, $p < .001$). Females were more likely to have the cannabis dependence symptom of physical–psychological problems (MH odds ratio = 3.1, $p < .001$).

TICs

For both alcohol and cannabis, TICs showed a single marked peak, and test information dropped off markedly at both lower and higher levels of substance problem severity. Alcohol symptoms provided a test information peak that was lower than that for cannabis symptoms (5.6 vs. 7.3) and at a higher level of problem severity (peak TICs occurred at latent trait values of 0.20 for alcohol vs. −0.59 for cannabis). With regard to alcohol symptom count, average test information values were 2.6 (for those with 1 symptom), 4.0 (2 symptoms), 5.2 (3 symptoms), 5.6 (4 symptoms), 5.5 (5 symptoms), 5.1 (6 symptoms), 4.6 (7 symptoms), 3.9 (8

Reliability, Validity, Test Theory (Psychology 694a/621)

I. Reliability
   A. Assessed by the reliability coefficient r_{tt}
   B. Methods of estimating r_{tt}

Student\ Item	1	2	3	4	5	6	T	T' (parallel)	T'' (retest)
1	1	1	1	1	0	0	4	5	5
2	1	1	1	1	1	1	6	5	6
3	1	1	1	0	0	0	3	3	4
4	0	0	0	0	1	0	1	0	1
5	0	0	0	0	0	1	1	2	1

1. Parallel forms -- if two forms are truly interchangeable

Student	T	T'	T^2	T'^2	TT'
1	4	5	16	25	20
2	6	5	36	25	30
3	3	3	9	9	9
4	1	0	1	0	0
5	1	2	1	4	2
	15	15	63	63	61

NOTE: In all following formula, N = # examinees, n = # items

Computational formula for correlation

$$r_{TT'} = \frac{N\Sigma T'T - (\Sigma T)(\Sigma T')}{\sqrt{(N\Sigma T^2 - (\Sigma T)^2)(N\Sigma T'^2 - (\Sigma T')^2)}}$$

Two kinds of variations in this design:
1. Day to day variations in people
2. Variations in set of items

$$r_{TT'} = \frac{(5)(61) - (15)(15)}{\sqrt{((5)(63) - (15)^2)((5)(63) - (15)^2)}}$$

Can therefore determine how resistant test is to both these kinds of variations.

$$r_{TT'} = \frac{305 - 225}{\sqrt{(315 - 225)(315 - 225)}} = \frac{80}{90} = .89$$

2. Test-retest

Student	T	T"	T^2	$T"^2$	TT"
1	4	5	16	25	20
2	6	6	36	36	36
3	3	4	9	9	12
4	1	1	1	1	1
5	1	1	1	1	1
	15	17	63	79	70

$$r_{T"T} = \frac{N\Sigma T"T - (\Sigma T)(\Sigma T")}{\sqrt{(N\Sigma T^2 - (\Sigma T)^2)(N\Sigma T"^2 - (\Sigma T")^2)}}$$

$$r_{T"T} = \frac{(5)(70) - (15)(17)}{\sqrt{((5)(63) - (15)^2)((5)(79) - (17)^2)}} = \frac{95}{\sqrt{(90)(106)}} = .97$$

3. Internal consistency -- used for a single set of test scores
   a. Split-half reliability

Student	O ($i_1+i_3+i_5$)	E ($i_2+i_4+i_6$)	O^2	E^2	OE
1	2	2	4	4	4
2	3	3	9	9	9
3	2	1	4	1	2
4	1	0	1	0	0
5	0	1	0	1	0
	8	7	18	15	15

$$r_{OE} = \frac{N\Sigma OE - (\Sigma O)(\Sigma E)}{\sqrt{(N\Sigma O^2 - (\Sigma O)^2)(N\Sigma E^2 - (\Sigma E)^2)}}$$

$$r_{OE} = \frac{(5)(15) - (8)(7)}{\sqrt{((5)(18) - 8^2)((5)(15) - 7^2)}} = \frac{75 - 56}{\sqrt{(26)(26)}} = .73$$

But, actual test is 6 items. Therefore need to correct this reliability estimate using the *Spearman-Brown Prophecy Formula*:

$$r_{TT_{new}} = \frac{mr_{TT_{old}}}{1+(m-1)r_{TT_{old}}} = \frac{(2)(.73)}{1+(1)(.73)} = \frac{1.46}{1.73} = .84$$

where: $m = \frac{n\ items \in NEW\ test}{n\ items \in OLD\ test}$

DO NOT USE SPLIT HALF for SPEEDED TESTS
Assume you have a long test with very easy items (***p*** . 1.0, therefore r_{it}.0.0), only difference between examinees is # completed

Student	O ($i_1+i_3+...+i_{n-1}$)	E ($i_2+i_4+...+i_n$)	O^2	E^2	OE
1	10	10	100	100	100
2	11	11	121	121	121
3	15	15	225	225	225
4	18	18	324	324	324
5	20	20	400	400	400
	74	74	1170	1170	1170

$$r_{OE} = \frac{(5)(1170)-(74)(74)}{\sqrt{((5)(1170)-74^2)((5)(1170)-74^2)}} = \frac{374}{\sqrt{(374)(374)}} = 1.0$$

Note Spearman-Brown will leave unchanged, 2*1 / 1+1 = 1.0

b. Kuder-Richardson Formula 20

Student / Item	1	2	3	4	5	6	T
1	1	1	1	1	0	0	4
2	1	1	1	1	1	1	6
3	1	1	1	0	0	0	3
4	0	0	0	0	1	0	1
5	0	0	0	0	0	1	1
p_i	.6	.6	.6	.4	.4	.4	ΣT=15
$1-p_i$	.4	.4	.4	.6	.6	.6	Mean=15/5=3
$p_i(1-p_i)$	.24	.24	.24	.24	.24	.24	σ_T^2=3.6

$$r_{KR20} = \frac{n}{n-1}\left[1 - \frac{\Sigma\, p_i(1-p_i)}{\sigma_T^2}\right]$$

$$r_{KR20} = \frac{6}{6-1}\left[1 - \frac{.24+.24+.24+.24+.24+.24}{3.6}\right] = .72$$

Conceptually, this is the mean of all possible split-half reliabilities, already corrected for double length.

c. Kuder-Richardson Formula 21-- Assumes that all items have the same ***p*** value. If all items do not have the same ***p*** value, then KR-21 is an underestimate.

Student	T	T^2
1	4	16
2	6	36
3	3	9
4	1	1
5	1	1
Total (Σ)	15	63

$$\overline{T} = \frac{15}{5} = 3$$

$$\sigma_T^2 = \frac{63}{5} - [\frac{15}{5}]^2 = 3.6$$

$$r_{KR21} = \frac{n\sigma_T^2 - \overline{T}(n - \overline{T})}{(n-1)\sigma_T^2}$$

$$r_{KR21} = \frac{(6)(3.6) - (3)(6\ \ 3)}{(6-1)(3.6)} = .70$$ Less than that obtained using KR-20 unless all $\boldsymbol{p_i}$ are the same.

d. Cronbach's α -- for items not scored 0/1 (e.g., Likert items)

Item 6 Student	1	2	3	4	5	6	T
1	1	1	1	1	0	0	4
2	1	1	1	1	1	1	6
3	1	1	1	0	0	0	3
4	0	0	0	0	1	0	1
5	0	0	0	0	0	1	1
	3	3	3	2	2	2	15

$$\sigma_T^2 = \frac{63}{5} - [\frac{15}{5}]^2 = 3.6$$

$$\sigma_{i_1}^2 = \sigma_{i_2}^2 = \sigma_{i_3}^2 = \frac{1^2 + 1^2 + 1^2 + 0^2 + 0^2}{5} - [\frac{3}{5}]^2 = .24$$

$$\sigma_{i_4}^2 = \sigma_{i_5}^2 = \sigma_{i_6}^2 = \frac{1^2 + 1^2 + 0^2 + 0^2 + 0^2}{5} - [\frac{2}{5}]^2 = .24$$

$$\alpha = \frac{n}{n-1}[\,1 - \frac{\Sigma\sigma_i^2}{\sigma_T^2}\,]$$

$$\alpha = \frac{6}{6-1}[\,1 - \frac{(6)(.24)}{3.6}\,] = .72$$

Note that this is the same as KR-20 when items are scored 0/1. this is true because $\sigma_i^2 = p_i q_i$ if items are scored dichotomously

e. Hoyt's reliability coefficient

Item 6 Student	i_1	i_2	i_3	i_4	i_5	i_6	T
1	1	1	1	1	0	0	4
2	1	1	1	1	1	1	6
3	1	1	1	0	0	0	3
4	0	0	0	0	1	0	1
5	0	0	0	0	0	1	1
	3	3	3	2	2	2	15

Follows an ANOVA Model:

N randomly selected people (***N***=5)
I randomly selected items (***I***=6)
r replications per cell (***r***=1)

$$SS_{total} = \sum_{1}^{N}\sum_{1}^{I}\sum_{1}^{r} i^2 - \frac{[\sum_{1}^{N}\sum_{1}^{I}\sum_{1}^{r} i]^2}{NIr} = 15 - \frac{15^2}{(5)(6)(1)} = 7.5$$

$$SS_{persons} = \frac{\sum_{1}^{N}[\sum_{1}^{I}\sum_{1}^{r} i]^2}{Ir} - \frac{[\sum_{1}^{N}\sum_{1}^{I}\sum_{1}^{r} i]^2}{NIr} = \frac{4^2 + 6^2 + 3^2 + 1^2 + 1^2}{(6)(1)} - \frac{15^2}{(5)(6)(1)} = 3$$

$$SS_{items} = \frac{\sum_{1}^{I}[\sum_{1}^{N}\sum_{1}^{r} i]^2}{Nr} - \frac{[\sum_{1}^{N}\sum_{1}^{I}\sum_{1}^{r} i]^2}{NIr} = \frac{3^2 + 3^2 + 3^2 + 2^2 + 2^2 + 2^2}{(5)(1)} - \frac{15^2}{(5)(6)(1)} = 0.3$$

$$SS_{withincell} = \sum_{1}^{N}\sum_{1}^{I}\sum_{1}^{r} i^2 - \frac{\sum_{1}^{N}\sum_{1}^{I}[\sum_{1}^{r} i]^2}{r} = 15^2 - \frac{15^2}{(1)} = 0$$

$$SS_{interaction} = SS_{total} - SS_{persons} - SS_{items} - SS_{withincell}$$

Hoyt's Coefficient = $1 - \frac{MS_{interaction}}{MS_{persons}}$

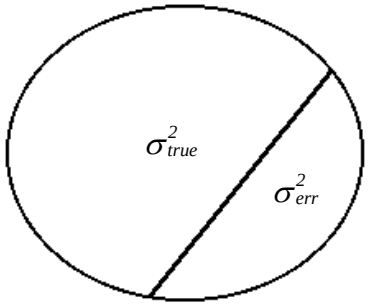
$$= 1 \frac{\frac{SS_{interaction}}{(N-1)(I-1)}}{\frac{SS_{persons}}{(N-1)}} = 1 \frac{\frac{4.2}{(5-1)(6-1)}}{\frac{3}{(5-1)}} = .72$$

Note: Algebraically equivalent to other α formulas when $\boldsymbol{r} = 1$. The keen thing about this formula is that it can be applied when more than one administration is given, and can estimate the error variance due to items, persons, and administrations.

4. Interjudge -- to be discussed below
   a. Intraclass
   b. Kappa

C. Factors affecting reliability coefficient

$$\sigma_T^2 = \sigma_{true}^2 + \sigma_{err}^2$$



$$r_{tt} = \frac{\sigma_{true}^2}{\sigma_{true}^2 + \sigma_{err}^2}$$

Underlying principle is that increasing test variance will lead to increase reliability. Can increase test variance by: increasing # items, tweaking item difficulties (close to middle range), increasing $\boldsymbol{r_{it}}$, by testing examinees that have a wider range of abilities, and by altering test content

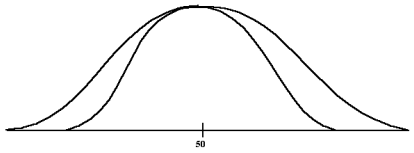
1. Number of items: $\uparrow \boldsymbol{n} \rightarrow \uparrow \boldsymbol{r_{tt}}$

$\boldsymbol{n}$	$\boldsymbol{r_{tt}}$
3	.73
6	.84
12	.91

2. Item difficulties (p_i values); medium p values enhance r_{tt}
3. Item discrimination (r_{it}); r_{it}68r_{tt}
4. Range of examinee abilities -- consider a 100-item mult choice test, with mean score = 50. *Heterogeneity is good*

   a. Group 1, σ^2 =100

$$r_{KR21} = \frac{n\sigma_T^2 - \bar{T}(n\bar{T})}{(n-1)\sigma_T^2} = .75$$

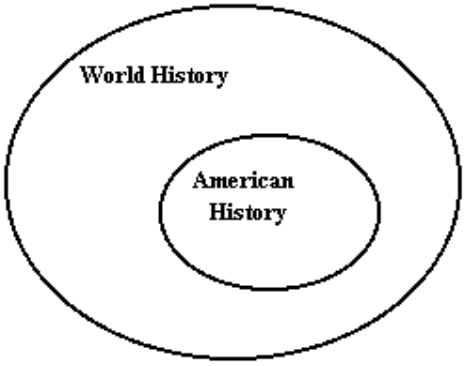


   b. Group 2, σ^2 =400

$$r_{KR21} = \frac{n\sigma_T^2 - \bar{T}(n\bar{T})}{(n-1)\sigma_T^2} = .95$$

5. Test Content -- *Homogeneity is good*

All else being equal (# items and mean difficulty), test on American history would have higher r_{tt}



   a. Item variance-Covariance Matrix:

	1	2	.....	n
1	s_1^2	s_{12}		s_{1n}
2	s_{21}	s_2^2		s_{2n}
...		...		
			...	
n	s_{n1}	s_{n2}		s_n^2

s_i^2 = variance of item i
s_{ij} = covariance of item i & j
= $s_i s_j r_{ij}$

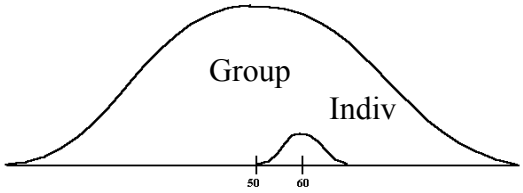
   b. σ^2 = sum of all entries in the table

c. If items more closely related to one another, increased covariance results and therefore σ^2 is larger as well. ($\uparrow \boldsymbol{\sigma^2} \rightarrow \uparrow \boldsymbol{r_{tt}}$)

6. Person reliability --

II. Standard Error of Measurement ($\boldsymbol{\sigma_e}$)

A. Consider 100-item multiple choice test, mean = 50, $\boldsymbol{\sigma_t}$ = 10, $\boldsymbol{r_{tt}}$ = .91



$\boldsymbol{\sigma_e} = \boldsymbol{\sigma_t}$ for individual across many administrations

$\boldsymbol{\sigma_e}$ = consistency of an individual's scores across many administrations

B. If T_{true} = subjects True test score, and T_{obs} = subject's observed test score, then T_{obs}- T_{true} = error; error is due to transient and irrelevant factors and may be either positive or negative

C. Therefore observed score T_{obs} contains some error and may overestimate or underestimate true score T_{true}.

D. $\boldsymbol{\sigma_e}$ can be estimated without repeated testings:

$$\sigma_e = \sigma_t \sqrt{1 - r_{tt}} = (10)(\sqrt{1 - 91}) = 3$$

Note: formula assumes that each examinee has same $\boldsymbol{\sigma_e}$; no way to overcome this assumption

E. Utility of $\boldsymbol{\sigma_e}$

1. Confidence Interval

$$T_{obs} - c\,\sigma_e \le T_t \le T_{obs} + c\,\sigma_e$$

$\boldsymbol{c}$ = 1, 68% CI, $\boldsymbol{c}$ = 2, 95% CI. $\boldsymbol{c}$ = 3, 99% CI

2. Assumes that distribution is normal, and that $\overline{T}_{obs} = T_{true}$

3. CI for difference between two scores on same test, $\boldsymbol{T_{obs1}}$= 50, $\boldsymbol{T_{obs2}}$= 40

$$(T_{obs1} - T_{obs2}) - c\sqrt{2}\,\sigma_e \le T_{t1} - T_{t2} \le (T_{obs1} - T_{obs2}) + c\sqrt{2}\,\sigma_e$$

If interval does not include 0, then it is (68%, 95%, 99%) probable that one individual is more able than another.

$$(50 - 40)\ 2\sqrt{2}\,(3) \le T_{t1} - T_{t2} \le (50 - 40) + 2\sqrt{2}\,(3)$$

$$10 - 8.49 \le T_{t1} - T_{t2} \le 10 + 8.49$$

$$1.51 \le T_{t1} - T_{t2} \le 18.49$$

F. In judging acceptability of σ_e , must consider the range of the entire distribution (e.g., 3 points versus 100 possible)

$$Goal: \frac{\sigma_e}{n} \leq .05, \text{ where } n = \text{total points possible}$$

Note that this ratio decreases as number of items increases

III. Reliability of Difference scores (Bad News...)

A. $D = T_1 - T_2$ (e.g., pretest-posttest design); T_1 scores have a reliability, so too do T_2 scores; reliability of D is a function of these separate reliabilities AND the intercorrelation between these measures

$$\sigma^2_{t_1} = 100; \sigma^2_{t_2} = 100; r_{t_1t_1} = .96; r_{t_2t_2} = .79; r_{t_1t_2} = .50$$

$$r_{dd} = \frac{r_{t_1t_1}\sigma^2_{t_1} + r_{t_2t_2}\sigma^2_{t_2} - 2r_{t_1t_2}\sigma_{t_1}\sigma_{t_2}}{\sigma^2_{t_1} + \sigma^2_{t_2} - 2r_{t_1t_2}\sigma_{t_1}\sigma_{t_2}}$$

$$r_{DD} = \frac{.96(100) + .79(100) - 2(.50)(10)(10)}{100 + 100 - 2(.50)(10)(10)} = .75$$

1. Note that r_{DD} is considerably lower than either $r_{t_1t_1}$; $r_{t_2t_2}$
2. To protect against unreliability of difference scores:
   a. Obtain reliable T_1 and T_2 scores
   b. Strive to have correlation between T_1 and T_2 scores low

IV. Correction for attenuation

A. Measurement error in two sets of scores attenuates the correlation between them -- only reliable variance can correlate

B. $r_{t_1t_2} = .80$ e.g., Stanford-Binet and WAIS; if knew the true rather than the observed scores,what would the correlation be?
Assume $r_{t_1t_1} = .95; r_{t_2t_2} = .95$

C. $r_{t_{1true}t_{2true}} = \frac{r_{t_1t_2}}{\sqrt{r_{t_1t_1}r_{t_2t_2}}} = \frac{.80}{\sqrt{(.95)(.95)}} = .84$

Can therefore conclude that Binet and WAIS are measuring different things since corrected correlation still < 1.0

D. Formula is often used to determine the extent to which two tests are measuring the same thing.

E. Formula can also give you an indication of how well variables would correlate if you improved your sloppy tests
Assume $r_{t_1t_1} = .75; r_{t_2t_2} = .70; r_{t_1t_2} = .30$

$$r_{t_{1true}t_{2true}} = \frac{r_{t_1t_2}}{\sqrt{r_{t_1t_1} r_{t_2t_2}}} = \frac{.30}{\sqrt{(.75)(.70)}} = .41$$

V. Interjudge Reliability

A. Kappa coefficient -- a chance-corrected measure of interjudge agreement for two judges and dichotomous classification.

$\kappa = \frac{p_o - p_c}{1 - p_c}$; $p_o =$ *proportion agreement obtained* ; $p_c =$ *proportion agreement by chance* ;

I9 Judge II6	Disease	NonDisease	
Disease	.75	.05	.80
NonDisease	.05	.15	.20
	.80	.20	

To compute $\boldsymbol{p_o}$, add up proportions where judges agree (diagonal). To compute $\boldsymbol{p_c}$, use marginal proportions; conceptually, what would be proportion agreement if you arbitrarily assign label of diseased to 80% of folks?

$p_o = H_t = .90$; $p_c = (.8)(.8) + (.2)(.2) = .68$

$$\kappa = \frac{p_o - p_c}{1 - p_c} = \frac{.90 - .68}{1 - .68} = .69$$

Kappa has been criticized that if base rates deviate markedly from 50%, Kappa will be low:

I9 Judge II6	Disease	NonDisease	
Disease	.04	.06	.10
NonDisease	.01	.89	.90
	.05	.95	

93% agreement -- but...

$p_o = H_t = .93$; $p_c = (.05)(.10) + (.90)(.95) = .86$

$\kappa = \frac{p_o - p_c}{1 - p_c} = \frac{.93 - .86}{1 - .86} = .50$ Pretty flimsy; Kappa definitely susceptible to base rates.

In general, $\kappa \geq .80$ is very good, $\geq .70$ acceptable, $< .50$ -- go back to drawing board

B. Intraclass correlation -- 2 or more judges, any type of data
   1. for example, consider ratings on a patient:

Patient #	Judge #1	Judge #2
1	3	4
2	5	6
3	7	7
...	10	9

   Whereas the Pearson product-moment correlation would be high (based on ***Z*** scores), the intraclass correlation would be lower because of disagreement over absolute values.
   2. Also good for assessing repeated testing. If scores change, intraclass correlation will reflect that -- will be lower
   3. Also good for more than 2 judges, and even when not all judges rate all patients! -- e.g., five judges, each with ratings on each subject, or some judges judge some subjects, some judge others; doesn't matter if unequal observations for different patients
   4. To compute

Patient	Judge A	Judge B	Judge C	
1	x	x	x	← Group 1
2	x	x	x	← Group 2
3	x	x	x	← Group 3
...				

$$R_{intraclass} = \frac{F-1}{F+1};\ F = \frac{\sigma^2_{betw\ groups}}{\sigma^2_{within\ groups}}$$

σ^2 obtained from simple ANOVA -- instead of judges serving as group and patients as observations, patients serve as groups and judges serve as observations-- Only for n=2 raters

alternatively, $R_{intraclass} = \frac{MS_B - MS_W}{MS_B + (n-1)MS_W}$

n = number of observers or raters

A Comment on the Reliability of Difference Scores and the Overall and Woodward Paradox

Significance tests of differences can be powerful even if the reliability of the difference scores is near zero (Overall & Woodward, 1975; Zimmerman, Williams, & Zumbo, 1993). The paradox pointed out by Overall and Woodward (1975) is that difference scores with zero reliability can in fact give rise to high power to detect a significant difference. The paradox is resolved when one considers that reliability of the difference scores depends on the existence of variance in the difference score that can reliably rank-order individuals in terms of the *magnitude* of their difference scores, but that the power to detect a difference involves assessing a *mean* difference between the two scores relative to the variance in this difference score. Thus if one constituent score (e.g. Left activity) were for every subject a constant *k* less than the other constituent score (e.g. Right activity), then there would be no variability in the difference scores, and no reliability. On the other hand, the mean difference score would be *k*, with no variance around that mean, allowing for a powerful statistical test that the mean difference is significantly different than zero, and that a statistically significant difference has been found. The pragmatic implications are that the reliability of difference scores if of little consequence if one wishes to test the significance of such a difference (e.g. to test that Right activity is greater than Left activity for the group as a whole), but the reliability of the difference score will be highly relevant when one is using the difference score to examine how individual differences in that difference score relate to other variables of interest (e.g. how individual differences in the asymmetry score relate to individual differences in BAS scores). In the latter case, the reliability of the difference score will impose constraints on the magnitude of the correlation that can be observed, as the maximum correlation that can be observed between two variables will be the square root of the product of the reliability of the two variables. Thus, because a sizable portion of the research examining frontal EEG asymmetry is concerned with the relationship of individual differences in frontal EEG asymmetry to other individual difference measures, the reliability of the asymmetry metric assumes great importance.

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin, 82*, 85-86.

Zimmerman, D. W., Williams, R. H., & Zumbo, B. D. (1993). Reliability of measurement and power of significance tests based on differences. *Applied Psychological Measurement, 17*, 1-9.

A bit more on Coefficient Alpha

Cortina, J. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, *78*, 98-104.

"An adequate coefficient alpha (number of items notwithstanding) suggests only that, on the average, split halves of the test are highly correlated. It says nothing about the extent to which the two halves are measuring the construct **or contructs** that they are intended to measure. Even if the total score of a test could perhaps be used for some practical purpose like selection, it could not be interpreted. In other words, the test would be known to measure something consistently, but what that is would still be unknown. Some form of construct validation is necessary to establish the meaning of the measure."

John's take home message:

- In other words, high internal consistency does not guarantee unidimensionality!
- High internal consistency does suggest few or no items that draw unique variance.

Which Intraclass Correlation is Right for You?!

Case 1. One has a pool of raters. For each subject, one randomly samples from the rater pool k different raters to rate this subject. Therefore the raters who rate one subject are not necessarily the same as those who rate another. This design corresponds to a 1-way Analysis of Variance (ANOVA) in which Subject is a random effect, and Rater is viewed as measurement error.

- ICC(1,1): used when each subject is rated by multiple raters, raters assumed to be randomly assigned to subjects, all subjects have the same number of raters.

- ICC(1,k): Same assumptions as ICC(1,1) but reliability for the mean of k ratings.

Case 2. The same set of k raters rate each subject. This corresponds to a *fully-crossed* (Rater × Subject), 2-way ANOVA design in which both Subject and Rater are separate effects. In Case 2, Rater is considered a *random* effect; this means the k raters in the study are considered a random sample from a population of potential raters. The Case 2 ICC estimates the reliability of the larger population of raters.

- ICC(2,1): used when all subjects are rated by the same raters who are assumed to be a random subset of all possible raters.

- ICC(2,k): Same assumptions as ICC(2,1) but reliability for the mean of k ratings.\

Case 3. This is like Case 2--a fully-crossed, 2-way ANOVA design. But here one estimates the ICC that applies only to the k raters in the study. Since this does not permit generalization to other raters, the Case 3 ICC is not often used.

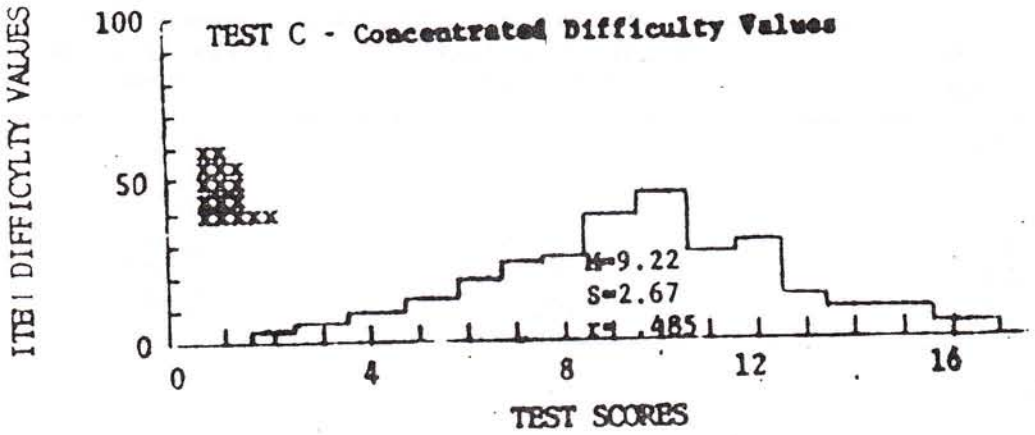
- ICC(3,1): used when all subjects are rated by the same raters who are assumed to be the entire population of raters.

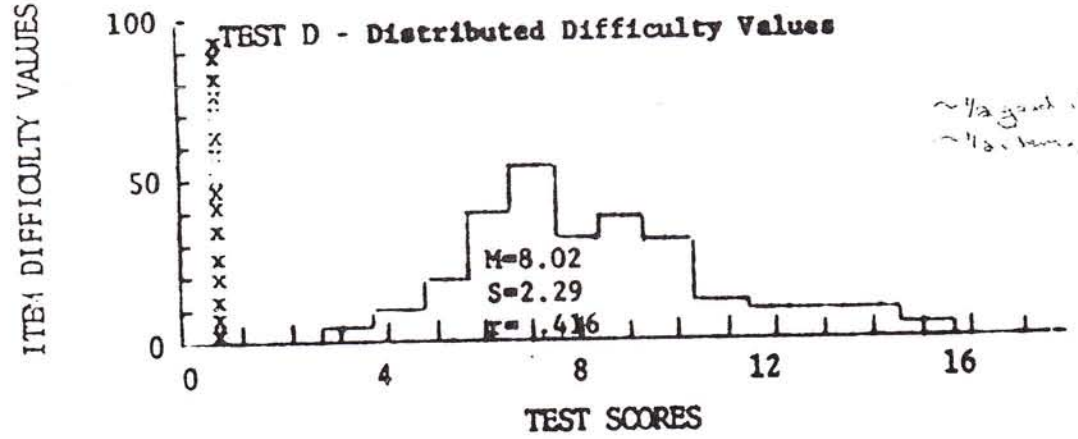
- ICC(3,k): Same assumptions as ICC(3,1) but reliability for the mean of k ratings. Assumes additionally no subject by judges interaction.

Useful Elaborations:

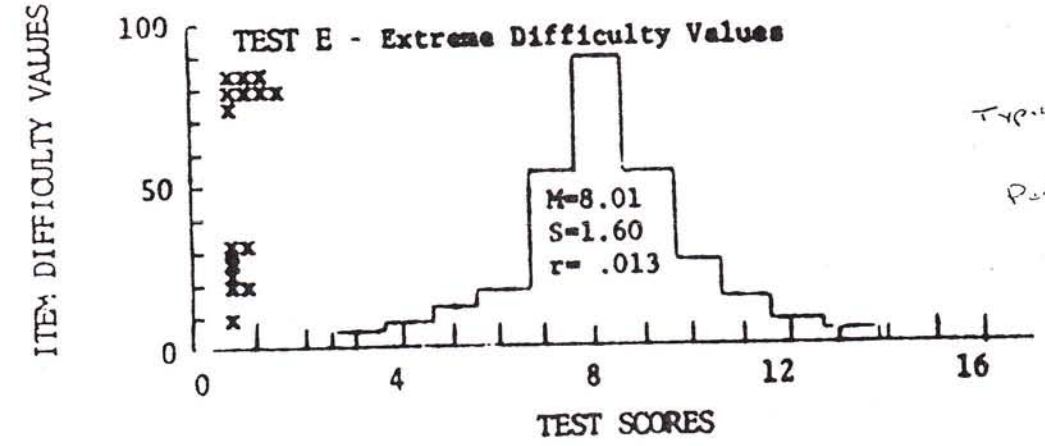
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.
- Intraclass Correlations with SPSS: http://www.nyu.edu/its/socsci/Docs/intracls.html

10/17/85





~½ good items
~½ items too hard or easy



Typical student gets
8 right, 8 wrong
Performance more consistent
no guessing

FIGURE 14.1
Relation of Distribution of Test Scores to Distribution of Item Difficulty Values

TABLE 14.4. Relation of Item Discrimination to Test Reliability For a One Hundred-Item Test

Mean Index of Discrimination	Standard Deviation of Scores	Reliability of Scores
0.1225	5.0	0.00
0.16	6.53	0.42
0.20	8.16	0.63
0.30	12.25	0.84
0.40	16.32	0.915
0.50	20.40	0.949

Reliability, Validity, Test Theory (Psychology 694a/621) continued...

VI.Validity

A.Overview:

B.Three non-mutually-exclusive types of validity

1.Content Validity -- Are items a representative sample of the content domain or universe of items that may be asked?

2.Criterion Validity

a.involved when test is used to estimate or predict behavior ***external*** to the measuring instrument; i.e., a criterion!

(1)Determined by the size of the correlation between the test instrument and criterion -- bigger ***r*** is better (absolute value)

b.Three types of criterion validity

(1)Predictive

(2)Concurrent

(3)Postdictive

c.Methods

(1)Contrasted groups (***t***-test method)

(2)Correlate test scores with behaviors

(3)Correlate test scores with other tests that are purported to measure something similar;

d.Methodological considerations:

(1)Correlation between two measures (e.g. predictor and criterion) is limited by restricted range

(a)Correction for restricted range in predictor variable:

$$r_{xy}^2 = \frac{\lambda r_{\tilde{x}\tilde{y}}^2}{1+(\lambda-1)r_{\tilde{x}\tilde{y}}^2}; \lambda = \frac{\sigma_x^2}{\sigma_{\tilde{x}}^2} = \frac{144}{36} = 4$$

$$r_{xy}^2 = \frac{(4)(.20)}{1+(4-1)(.20)} = .50, \therefore r = .71$$

(2)Correlation between two measures (e.g. predictor and criterion) is limited by the reliability of each:

$$r_{t_1t_2} \le \sqrt{r_{t_1t_1}}\ ; r_{t_1t_2} \le \sqrt{r_{t_2t_2}}$$

Correction for attentuation is based on this principle

$$r_{t_{1true}t_{2true}} = \frac{r_{t_1t_2}}{\sqrt{r_{t_1t_1}r_{t_2t_2}}}$$

Often, investigators go to great lengths to ensure reliability of their predictor instruments, but criterion variables may or may not be as reliably measured.

(3)Standard error of estimate indicates how specific your prediction of the criterion scores is

$$\sigma_{est} = \sigma_{criterion}\sqrt{1 - r_{xy}^2} \qquad r = .80, \sqrt{1-r^2} = .60$$

In this case, error is 60% as large as if guessing (i.e., mean).

3.Construct Validity

a.Construct ≡ attribute for which it is often difficult to develop an operational definition

b.Construct Validity ≡ Does your test measure the construct you purport -- and not other constructs

(1)Relevant traits

(2)Irrelevant traits

c.The Process of establishing construct validity

(1)Begin with a vague concept or construct
(2)Generate or evolve a theory surrounding your construct
(a)This process results in a theory: an interlocking system of laws that relate consturcts to one another and to tangible behaviors
(b)This interlocking system is *AKA* as a *nomological net*
(c)The process, schematically:
(d)Bootstrapping: *"Intelligence is what the tests test"* -- BORING (1922)

d.Common methods for determining construct validity
(1)Correlational studies
(2)Factor Analysis -- consider the ideal matrix below:

Test	Math	Spelling	Jumping	Running
Math	1.0	.81	0	0
Spelling	.81	1.0	0	0
Jumping	0	0	1.0	.81
Running	0	0	.81	1.0

Submit the matrix to a factor analysis. Factor analysis will produce another matrix that accounts for most of the original σ^2 with fewer factors than the original number of variables (or tests). This is the factor loading matrix, which summarizes the intercorrelation between the original variables (tests) and new hypothetical varibles labelled Factor I and Factor II:

Test↓ / Factor→	I	II
Math	.90	0
Spelling	.90	0
Jumping	0	.90
Running	0	.90

In this example, two hypothetical factors are determining performance; Factor I is Cognitive Abilities and is responsible for performance on Math and Spelling tests; Factor II is Physical Abilities and is responsible for performance on Jumping and Running tests.
(a)Labelling the factors is necessarily subjective.
(b)Method above uses several tests (*ala* Campbell and Fiske); can also subject items to factor analysis to see if more than one construct may be accounting for your test σ^2
(3)Experimental attempts to alter scores on a test -- certain manipulations should alter test scores, others should not
(a)e.g., Scores on WAIS-R should be resistant to training if they are a true measure of ability; of course, training with very similar items may increase scores, which would demonstrate what we all know -- that in addition to general intellectual abiliity, the WAIS-R taps item-specific abilities