

# snr\_\_wordspan

*Christopher W Bishop*

*10/18/2015*

## Contents

Overview . . . . .	1
Recognition . . . . .	3
Judgment . . . . .	3
Recall . . . . .	4

## Overview

Bishop and Miller talked today and Miller mentioned that there may be intersite differences between UW and U of I in Word Span scores. Bishop and Miller reviewed data provided in XLSX spreadsheet via e-mail. Bishop wanted to confirm Miller's observations.

```
# Load libraries
```

```
library(readr)
library(stringr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(foreach)
library(ggplot2)
```

```
# Load data
```

```
data = read_csv("HASNR_DATA_2015-10-18_1020.csv")
```

```
# Get site information
```

```
site = unlist((data %>% rowwise() %>% do(site = if (.$subject_id>1999){return('U of I')} else{'UW'})))$s
```

```
# Add to data
```

```
data = data %>% mutate(site = site)
```

```
# Filter so we are only looking at wordspan information
```

```
# RedCap data frame is very big and very confusing unless filtered down
```

```

wordspan = data %>% select(subject_id, contains('wordspan'))

# Concatenate scores and add set size information
# Do for recognition, judgment, recall
recognition = wordspan %>% select(subject_id, contains('recognition'))

reshape.wordspan <- function(df_table, test_segment){
  "
  Reshapes REDCAP word span data into a more useful format.
  "

  # Filter df_table to only include wordspan and test_segment
  df_table = df_table %>% select(site, subject_id, contains(paste0('wordspan_', test_segment)))

  # Get all names of relevant columns
  cnames = names(df_table %>% select(contains(test_segment)))

  # For each column name, create a data frame with subject ID, test_segment, and set size
  wordspan <- foreach (cname=cnames, .combine=rbind) %do% {

    # Temporary table containing the selected data
    tmp_table = df_table %>% select(site, subject_id, matches(cname))

    set_size = unlist(str_split(cname, '_'))[3]

    # Add set size
    tmp_table = tmp_table %>% mutate(set_size=as.integer(set_size))

    # Rename to test segment
    tmp_table = tmp_table %>% rename_(test_segment = cname)

    return(tmp_table)
  }

  names(wordspan)[3] <- test_segment
  return(wordspan)
}

# Get test components.
recognition = reshape.wordspan(data, 'recognition')
recall = reshape.wordspan(data, 'recall')
judgment = reshape.wordspan(data, 'judgment')

# Join them segments together
wordspan = recognition %>% left_join(recall, by=c('site', 'subject_id', 'set_size')) %>% left_join(judgment, by=c('site', 'subject_id', 'set_size'))

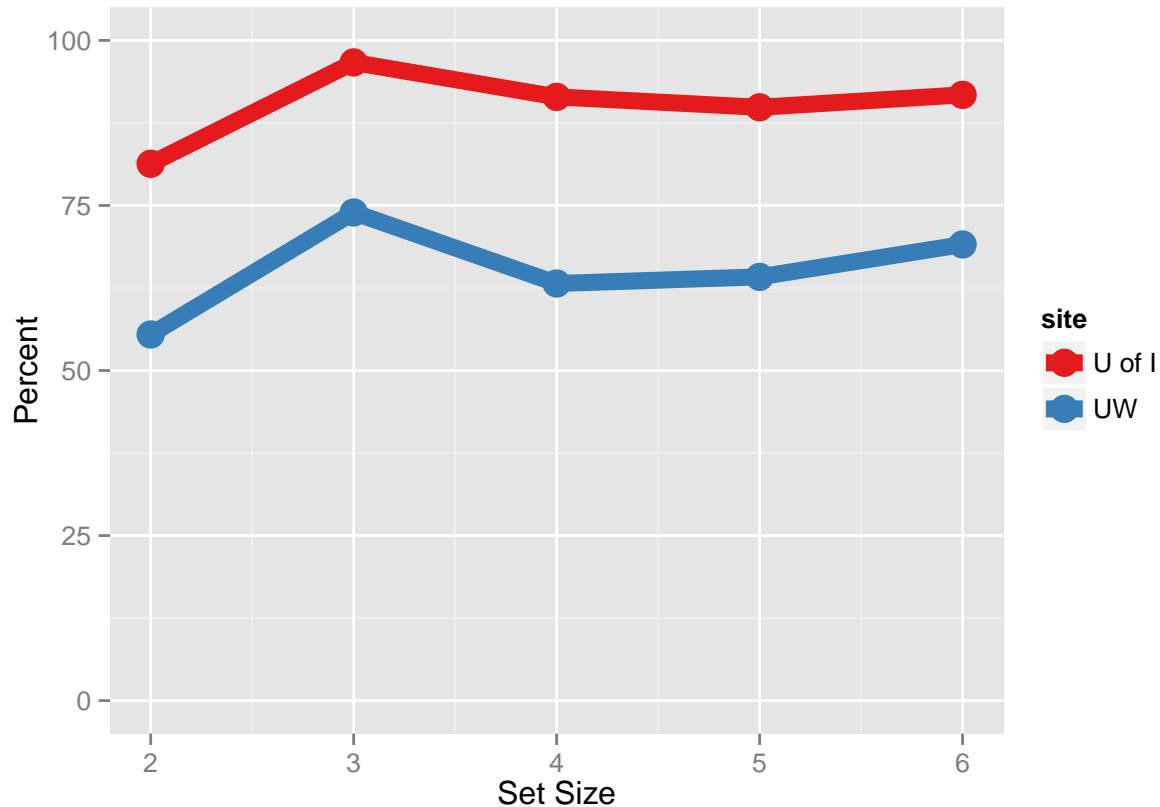
# Summary statistics
wordspan_summary = wordspan %>% group_by(site, set_size) %>% summarise_each(funs(mean(., na.rm=TRUE)))

```

## Recognition

```
# Plot recognition
```

```
ggplot(wordspan_summary, aes(x=set_size, y=recognition, colour=site)) + geom_point(size=5) + geom_line(size=2)
```



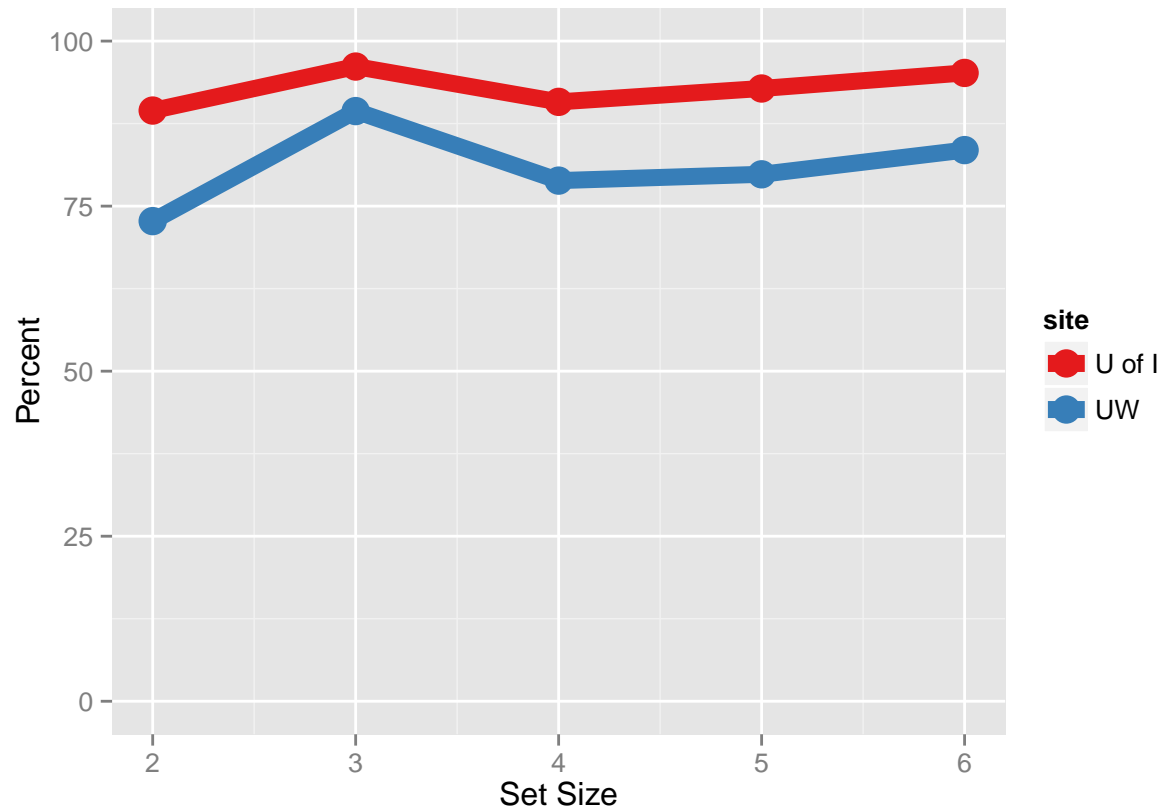
There are two clear trends in the data.

- UW subjects are performing poorer overall compared to U of I.
- Subjects generally do poorly in set size of two.
  - This is likely a residual training effect. From Bishop's few and informal observations, it seems that subjects do not always understand instructions well.
  - Practice sessions might be informative here.
- It looks like people do *best* with set size of 3, which is strange.
  - Need to dig into this more.

## Judgment

```
# Plot recognition
```

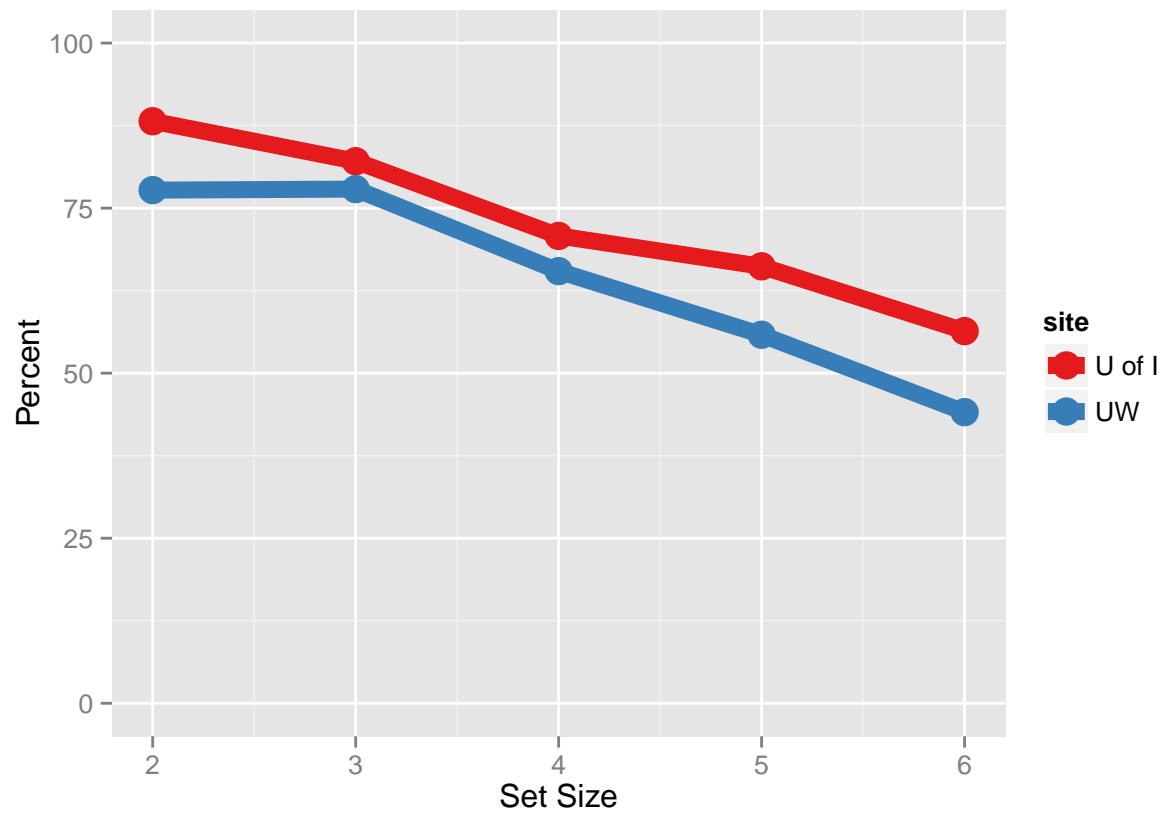
```
ggplot(wordspan_summary, aes(x=set_size, y=judgment, colour=site)) + geom_point(size=5) + geom_line(size=2)
```



Similar trends here. Not sure what to make of the bump at set size of 3.

## Recall

```
# Plot recognition
ggplot(wordspan_summary, aes(x=set_size, y=recall, colour=site)) + geom_point(size=5) + geom_line(size=1)
```



- Again, UW subjects are doing slightly worse.
- The performance is nearly parallel.
- Bishop really needs to break this down at individual level. No time to do so right now.